Library

NOV 5 1946

Vol. 36, No. 5 October, 1946

# Journal of Experimental Psychology

EDITED BY

SAMUEL W. FERNBERGER, ON LEAVE
FRANCIS W. IRWIN, ACTING EDITOR
UNIVERSITY OF PENNSYLVANIA

CONSULTING EDITORS

CHARLES W. BRAY, *Princeton University;* ELMER K. CULLER, *University of Rochester;* CLARENCE H. GRAHAM, *Brown University;* JOY P. GUILFORD, *University of Southern California;* FRANCIS W. IRWIN, *University of Pennsylvania;* DONALD G. MARQUIS, *Yale University;* ARTHUR W. MELTON, *University of Missouri;* CARROLL C. PRATT, *Rutgers University.*

## CONTENTS

*The Role of Secondary Reinforcement in a Partial Reinforcement Learning Situation:* M. R. DENNY ..... 373

*Dark Adaptation as a Factor in the Sensitization of the Beta Response of the Eyelid to Light:* D. A. GRANT AND EUGENIA B. NORRIS ..... 390

*Level of Aspiration: Ambition or Defense?:* R. B. HOLT ..... 398

*The Effect of Context upon Judgments of Pitch Differences:* T. KOESTER AND W. N. SCHOENFELD 417

*Development of Speech during Infancy: Curve of Phonemic Types:* O. C. IRWIN AND H. P. CHEN 431

*Adaptation in Energy Mobilization: Changes in General Level of Palmar Skin Conductance:* ELIZABETH DUFFY AND O. L. LACEY ..... 437

*Validity of Frequency of Blinking as a Criterion of Readability:* M. A. TINKER 453

*The Effects of Certain Pre-Training Procedures upon Maze Performance and Their Significance for the Concept of Latent Learning:* H. W. KARN AND J. M. PORTER, JR. 461

PUBLISHED BI-MONTHLY
BY THE AMERICAN PSYCHOLOGICAL ASSOCIATION, INC.
PRINCE AND LEMON STS., LANCASTER, PA.
AND 1515 MASSACHUSETTS AVE., N. W., WASHINGTON 5, D. C.

Entered as second-class matter, February 6, 1937, at the post office at Lancaster, Pa. under the Act of March 3, 1879

Journal of
Experimental Psychology

VOL. 36, NO. 5 OCTOBER, 1946

# THE ROLE OF SECONDARY REINFORCEMENT IN A PARTIAL REINFORCEMENT LEARNING SITUATION*

BY M. RAY DENNY[1]

*From Psychological Laboratory, University of Iowa*

## INTRODUCTION

It is generally agreed by psychologists that the strength of a response tendency or the likelihood of its occurrence in a situation is, among other variables, a function of the number ($N$) of occasions that the response has been associated in the past with the situation. Learning theorists disagree, however, in the specification of this variable, $N$. One group defines it in terms of the occurrence of a goal or reinforcing situation which in turn requires the presence of some motivating or drive state. According to this group, the principle of reinforcement or effect is one of the most important principles of learning.

Other psychologists take an opposite view. They deny the necessity of specifying the variable, $N$, in terms of goal attainments, reinforcements or after-effects. Reward of a response is not, according to them, a necessary condition for strengthening it. A goal is considered to play only a motivational or expectancy role. According to proponents of this view, a sufficient condition for learning is the occurrence of the response in contiguity with the situation; the greater the number of such associations the stronger will be the tendency to make the response.

One type of evidence that has led to a questioning of the point of view that the relevant variable determining response strength is the number of reinforcements, is to be found in studies in which

* This article is a major portion of a dissertation presented to the faculty of the Department of Psychology of the State University of Iowa in partial fulfillment of the requirements for the Ph.D. degree. The writer is indebted to Professor Kenneth W. Spence for his valuable advice and assistance throughout the course of this research.

[1] Now at Michigan State College.

partial reinforcement rather than the more usual condition of 100 percent reinforcement is used. Thus, a number of studies have been cited by proponents of the contiguity theory which show (1) that acquisition proceeds in much the same manner with partial reinforcement as with 100 percent reinforcement, and (2) that resistance of a learned response to extinction is actually greater following partial reinforcement than following 100 percent reinforcement. For example, Pavlov discovered in an exploratory study that conditioned responses developed rapidly if reinforcement was given only every second or third trial, although they did fail to develop with only 25 percent reinforcement. More recently Brogden (1), Cole (4), Brunswik (2), Skinner (12) and Humphreys (9, 10, 11) have found a high level of learning in both conditioning and instrumental situations when a partial reinforcement procedure was employed; and Skinner (12) and Humphreys (9, 10) have demonstrated greater resistance to extinction with partial than with 100 percent reinforcement. Such evidence has been accepted by Humphreys not only as refuting the reinforcement interpretation, but also as evidence that the relevant variable determining the strength of a learned response is the number of trials the response is practiced.

As an example of the nature of this evidence and as a point of departure for the present investigation the results of one of Humphreys' studies (11) may be cited in more detail. Using a Skinner-box set-up, he found that the presence of additional, randomly interspersed non-reinforced trials in the learning of a bar-pressing response by white rats resulted in significantly superior learning as measured by the number of responses in a 10-min. extinction period. All of his Ss were first given the same amount and kind of preliminary training. They were then divided into four groups and trained to press the lever under the four conditions of reinforcement which are presented, together with results, in Table I. The data given are the mean number of responses made in three extinction periods.

It is evident that fairly large differences occurred between groups 1 and 2 and between groups 3 and 4 but not between 2 and 3. Since the large differences occurred when the groups were matched for

TABLE I

HUMPHREYS' SKINNER-BOX DATA

| Group | No. Trials | No. Food Reinforcements | Mean No. Responses |
|---|---|---|---|
| 1 | 52 | 18 | 35.0 |
| 2 | 18 | 18 | 25.2 |
| 3 | 18 | 7 | 24.4 |
| 4 | 7 | 7 | 17.1 |

number of reinforcements but not for trials, and little or no difference appeared when the groups were matched for number of trials but not for reinforcements, Humphreys concluded that the important variable determining the strength of conditioning was the number of practice trials and not the number of reinforcements.

So far as the phenomenon of greater resistance to extinction of a conditioned response based on partial reinforcement is concerned, Hull (7) has suggested a plausible explanation in terms of certain conditioning principles. Hull considers it to be a stimulus patterning phenomenon, explainable in terms of the principle of perseverative receptor discharges and one or two related principles. As Hull points out, with 50 percent reinforcement the response that becomes conditioned is, in the main, conditioned to the after-effects of non-reinforcement, for in Humphreys' study 67 percent of the reinforcements follow non-reinforcement. In the 100 percent reinforcement group, on the other hand, the conditioned stimulus trace, on every occasion but the first, contains the after-effects of reinforcement. In the extinction series, however, the conditioned stimulus trace contains no after-effects of reinforcement, with the result that the stimulus trace for the 50 percent group during extinction is more similar to its counterpart during conditioning than is true with the 100 percent group. Consequently, the strength of the response reinforced 50 percent of the time is less weakened by factors of stimulus generalization and afferent neural interaction,[2] and hence is more resistant to extinction.

There is, however, still another factor that seems to have been overlooked by Humphreys in his interpretation of these experiments. This is the principle of secondary reinforcement.[3] It is entirely possible that secondary reinforcement was operative in the partial reinforcement groups on those trials in which the unconditioned or reinforcing stimulus was omitted. For it has been shown by Bugelski (3) and Skinner (12) that the click accompanying the movement of the lever just preceding the food reward in the Skinner box can acquire sub-goal or secondary reinforcing properties. Considering that Humphreys gave all his Ss 100 preliminary training trials in which a buzzer was followed immediately by food, it is not unreason-

[2] For Hull the principles of stimulus generalization and afferent neural interaction mediate the phenomenon of stimulus patterning; the effect of afferent neural interaction is to weaken the habit strength to the original stimulus complex. *Cf.* Hull (8), pp. 356–368.

[3] According to Hull (8), secondary reinforcement is the reinforcing power acquired by stimuli which have been closely and consistently associated with a reinforcing state of affairs. This principle is basic to Hull's gradient of reinforcement hypothesis by which he accounts for compound trial and error learning involving a protracted behavior sequence. In fact the extent of the reinforcement gradient is a function of the conditions providing for or precluding secondary reinforcement.

able to assume that the buzzer served as a secondary reinforcing agent, thus providing reinforcement on what were ostensibly non-reinforced trials.[4]

On the basis of this analysis it is presumed that the experimental finding of approximately equal amounts of learning with either partial or consistent reinforcement is due to the operation of secondary reinforcement. If experimental conditions could be arranged so that secondary reinforcement was not present on the non-reinforced (i.e., non-food) trials it would be possible to test this hypothesis: the inference is that under such conditions a difference would be obtained in the amount of learning under partial and full reinforcement.

The present study has attempted to compare learning with full and partial reinforcement under two sets of conditions—one in which no attempt has been made to control secondary reinforcement, and one in which secondary reinforcement, or at least its differential effect, has been greatly minimized. According to the present hypothesis, one would expect to find little difference between the full and partial reinforcement groups under conditions in which secondary reinforcement is uncontrolled, whereas a significant difference should occur when secondary reinforcement is controlled.

## Experimental Procedure

*A. Apparatus.* The apparatus used was a simple T-maze. The ground plan is presented in Fig. 1. The maze consisted of one pair of arms or delay boxes, four end or goal boxes, a com-

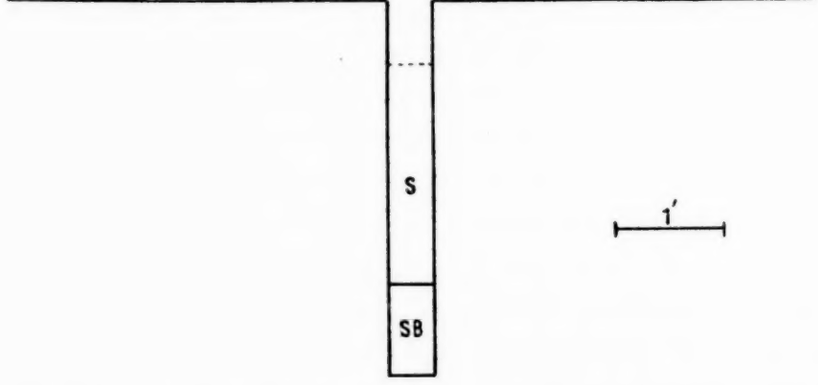


Fig. 1. Ground plan of apparatus. SB—starting boxes; S—stem; DB1, DB2—delay boxes C—curtains; D—doors; G—end boxes.

bined starting box and stem and a constant choice point section. The apparatus was movable and similar units were interchangeable. The sides and other wooden portions were constructed of white pine. The floors and roofs of the delay boxes and choice point were made of opal flashed

[4] On the assumption that secondary goal stimuli are not so effective as primary reinforcement we should expect the curves of acquisition under partial reinforcement to be slightly lower than those for 100 percent reinforcement. Such data on acquisition as exist, Humphreys (9) and Brunswik (2), lend support to this implication.

glass in order to provide a uniform perceptual field and to eliminate extra-maze visual cues. The starting box and the first 24 in. of the stem were fully enclosed in wood.

Opal flashed glass doors, 4½ × 6 in., were located at the exit of the starting box, in the choice point at both entrances to the stem and at the entrance to each delay box. The doors in the delay boxes at the entrance into the goal or end boxes were made of wood and painted grey, as were all the wood portions of the maze. The doors were operated by a system of strings and weights.

The sections of the maze were elevated in order to permit lighting from below. Small 7½ watt bulbs were installed on a wooden base a distance of 12 in. beneath the floor of the maze, a distance sufficient to give uniform lighting within the maze. Each light base was firmly attached to the sides of a maze section by wooden posts, providing a stable and easily movable unit that stood on the floor. This method of lighting permitted the *E* to follow the progress of the rat through the enclosed maze by means of its shadow.

Two of the end boxes were constructed so as to be as similar as possible, the same width, height and color as the alleys and with wood floor and opal glass roof. A third or positive goal box was made different from these; it was two in. wider and one in. higher than the alleys, was painted white, and had an opal glass roof and a beaverboard floor covered with hardware cloth. The fourth or negative end box was the same size and shape as the positive goal box, but it was painted black and the floor was not covered with hardware cloth. Suspended above the roof of each end box was a small 7½ watt light providing illumination for the feeding rat. The constant choice point section extended 10 in. on either side of the stem, and gray curtains hung eight in. from the choice point within this structure.

In the preliminary training a 52-in. straight alley runway was used. The first section, 28 in. in length, was constructed all of wood, while the remaining 24 in. had an opal glass floor and roof, a gray curtain, and the same kind of lighting as in the experimental T-maze.

*B. Subjects.*—The *S*s were albino and hooded rats from the rat colony of the psychology department of the State University of Iowa. The ages of the animals at the beginning of preliminary training ranged from 76 to 118 days (mean age = 92 days). Individual animals from each litter were randomly assigned to the various groups, so that an equal number of hooded and albino rats and approximately the same number of animals from each litter fell into each main sub-group. A total of 88 animals which had had no prior experimental experience were used. They were divided according to sex and breed as follows: Males, 56; females, 32; albino, 52; hooded, 36.

*C. Preliminary training.*—All animals were first given one week of habituation in which a strict food regimen was maintained. Each animal received eight gm. of Purina Dog Chow per day at the same hour in the morning as they were to run in the experiment. Each animal was then given three days of preliminary training in the straight alley maze.

On the first day pairs of *S*s were placed twice in the positive white goal box, which contained a food cup full of food, and were allowed to eat two to four pellets of food each. Following this, each rat received four reward trials on the straight alley, the reward consisting of one pellet of dog chow (0.20 gm.).

On the second day the animals were given six more trials, the third and fifth of which were non-reinforced. The black, negative end box was used upon such occasions, thus providing experience with non-reinforcement and minimizing to some degree generalization of secondary reinforcement to the black goal box. On reinforced trials the *S*s were removed from the goal box immediately after the pellet of food was eaten; on non-reinforced trials the animals were retained in the negative end box for 20 sec.

On the third day four more reinforced trials in the straight alley were given, making a total of 14 experiences with the positive goal box. Each *S* was then taken for the first time to the experimental T-maze and was given three free choice runs to determine its position preference. The neutral gray boxes were present at the ends of the T during these trials. The *S*s received no food and were retained in the end boxes for only 10 sec.

Finally, each *S* was given one forced trial to the side it chose most often during this series and was immediately rewarded by receiving food in the white goal box. This reinforcement was given with a view to compensating for the possible effects of the non-reinforcement to the preferred side. In the experiment proper all rats were trained to the path opposite that chosen most often during this preference series.

*D. Method in learning series.*—Immediately following the completion of preliminary train-

ing the *Ss* were placed at random in one of two groups of either the experimental condition or the control condition. Under the control condition there were two such groups, C–50 and C–100. Each group contained 20 animals. They were run under the more or less typical maze learning conditions, in that the various sections of the maze, including the goal boxes, remained constant. The goal box at the end of the correct alley was always the large white box with the hardware cloth floor, while the black box always appeared at the end of the incorrect alley. *Ss* in Group C–100 always received food reinforcement upon taking the designated correct alley while the *Ss* in Group C–50 were reinforced by food on only 50 percent of such choices. It was assumed that secondary reinforcement would be operative in this condition on those trials in which the *S* was not reinforced by food, for the animal entered the white goal box, the cues of which had been associated with food reinforcement in preliminary training and had presumably acquired secondary reinforcing properties. Furthermore, the fact that the delay chambers remained on the same side meant that any differential cues of the delay box on the correct side would become secondary goals.

All *Ss* were delayed 15 sec. in either delay box. The purpose of the delay was to slow down learning and, in the case of the experimental condition, to minimize the secondary reinforcing effects based on the choice point cues and the proprioceptive cues of the animal's response.

The experimental condition also had two groups, X–50 and X–100, with 24 animals in each. They were run under special experimental conditions which were designed to eliminate as much as possible the operation of secondary reinforcement of those trials to the correct side on which the 50 percent reinforcement group, X–50, did not obtain food. To accomplish this one of the neutral gray boxes instead of the white, positive goal box was used at the end of the correct alley on the trials that no food was given. It was assumed that stimulus generalization of secondary reinforcement from the white to the very different (in brightness, size, flooring, etc.) gray boxes would be negligible. Moreover, in order to minimize the possibility that cues from the delay boxes, any constant floor cues or cues from a constant spatial source, such as a constant noise, would acquire secondary reinforcing properties, the delay boxes were interchanged according to a prearranged order and the starting box, stem and choice point sections were faced in opposite directions (east or west) on successive days.[5]

The experimental design is summarized briefly in the following table:

TABLE II

Summary of Experimental Design

| Secondary Reinforcement Cues | Food Reinforcement | |
|---|---|---|
| | 50% | 100% |
| Present | C–50 | C–100 |
| Absent | X–50 | X–100 |

The details of the daily experimental routine were as follows: All the *Ss* receiving 50 percent reinforcement (groups C–50 and X–50) and half of each of the groups (C–100 and X–100) receiving 100 percent reinforcement were given four trials in the maze per day. The first two trials of the day permitted free choice of either alley, while the last two were forced in such a manner as to equalize the number of responses to the correct and wrong paths. The method of giving forced trials avoided the establishment of a simple alternation habit, so that only the following four patterns of runs were possible: RRLL, LLRR, RLLR, and LRRL. For the *Ss* that were reinforced on only 50 percent of the correct responses, half of the time the first correct response of the day was reinforced and half the time the second. Whether the first or second correct response was reinforced on any experimental day was determined by a prearranged schedule.

The other half of groups C–100 and X–100 was given six trials a day instead of four. The first two trials were again free, but two extra forced trials were given in which the *S* was required to take the wrong alley. These *Ss* thus went twice to the correct alley and four times to the wrong alley each day. The purpose of this procedure may be best understood by pointing to the fact that in the 50 percent reinforcement groups the number of non-reinforcements on the wrong path at any time in the learning was always twice the number of reinforcements on the correct

[5] It should be noted, however, that, aside from room floor cues, the cues of the choice point section remained as constant in the experimental condition as in the control.

path. In the 100 percent reinforcement groups, on the other hand, the number of such non-reinforcements on the incorrect side always equaled the number of reinforcements on the correct side. For example, in the 50 percent reinforcement groups at the point at which four reinforcements had been given to the correct side, there were eight non-reinforcements to the wrong side, whereas in the 100 percent reinforcement groups the number of non-reinforcements to the wrong path was only four, the same as the number of reinforcements to the correct path.

In other words, the addition of two forced trials to the wrong path in half of the 100 percent reinforcement *Ss* provided a better comparison with the 50 percent reinforcement groups in terms of equalizing the total number of non-reinforcements to the wrong alley. This technique also made possible a study of the effects of variation of the number of non-reinforcements of the wrong response under comparable conditions of reinforcement of the correct response.

The 100 percent reinforcement *Ss* were trained for a period of six days and the 50 percent animals for a period of 12 days. The *Ss* were run in blocks of from six to 14 animals. The same order of running was followed on successive trials of one day, but differed from day to day. Some *Ss* in a block were trained to the left alley and others to the right alley. Either all control or all experimental animals were run in a block, but both 50 percent and 100 percent groups were run together.

In the event that an animal refused to enter the end box of the wrong path 30 sec. after the door was opened, it was removed from the alley box. Refusal of an *S* on forced trials to enter a wrong alley also led to its removal from the maze after a period of 60 sec. Records were kept of all such incomplete responses.

The inter-trial interval was at least 20 min. and varied up to 30 min. This length of period minimized the tendency to alternate responses. All *Ss* were retained in the end box 20 sec. when no food was given. *E* also kept a record of the running times for each trial by means of a stop watch. The period measured was from the opening of the starting box door to the rat's entrance into the delay box.

Food reward consisted of two medium-sized pellets of Purina Dog Chow (0.40 gm.). At the completion of each day's run the *Ss* were fed eight gm. of dog chow in an individual feeding cage. Thus, a period of 21 to 22 hours of food deprivation preceded each day's experimental period.

*E. Method in extinction series.*—At the end of 12 food reinforcements all animals were given 12 extinction trials on the day following the last learning day. In the extinction period all the doors in the maze including the ones to the goal box were opened and remained open until an animal entered a goal or end box whereupon the door was shut. The animal was never forced in any manner and was never reinforced by food. The time taken for the animal to make an initial left or right response into a delay box and the total time taken to enter a goal box were recorded, except where the rat did not hesitate in the least in going from delay box to goal box, whereupon just the time taken to enter the goal box was recorded. An animal was considered to have extinguished if it refused to enter a goal box after three minutes in the maze. The responses recorded were the initial left or right responses and the goal box finally entered.

Half of the animals in each sub-group, chosen at random, were extinguished with the positive white goal box present at the end of the previously correct alley, and half were extinguished with the neutral gray box present in the former goal position. However, on the first extinction trial the positive white box was present for all animals.

During the extinction series the inter-trial interval was kept at approximately 40 sec. For the first 15 sec. of this interval the *S* was retained in the end box it had entered.

## Results

*A. Learning measures.*—The principal measures of learning employed are the number and percentage of correct responses on the initial free trial of each day. While similar measures are available for the first two free trials of each day, the former are considered superior for several reasons: (1) The first trial data provide for equalization of the number of previous food reinforcements and the number of non-reinforcements. (2) They are not affected by the

tendency toward spontaneous alternation which occurred despite the length of the inter-trial interval. (3) It is possible to use the initial trial of the extinction series as a final measure of learning.

The running time measures are not reported as they show essentially the same trends as the error measures.

*B. Comparison of learning under the control condition.*—The results for the two control groups are present in Fig. 2 and Table III. From

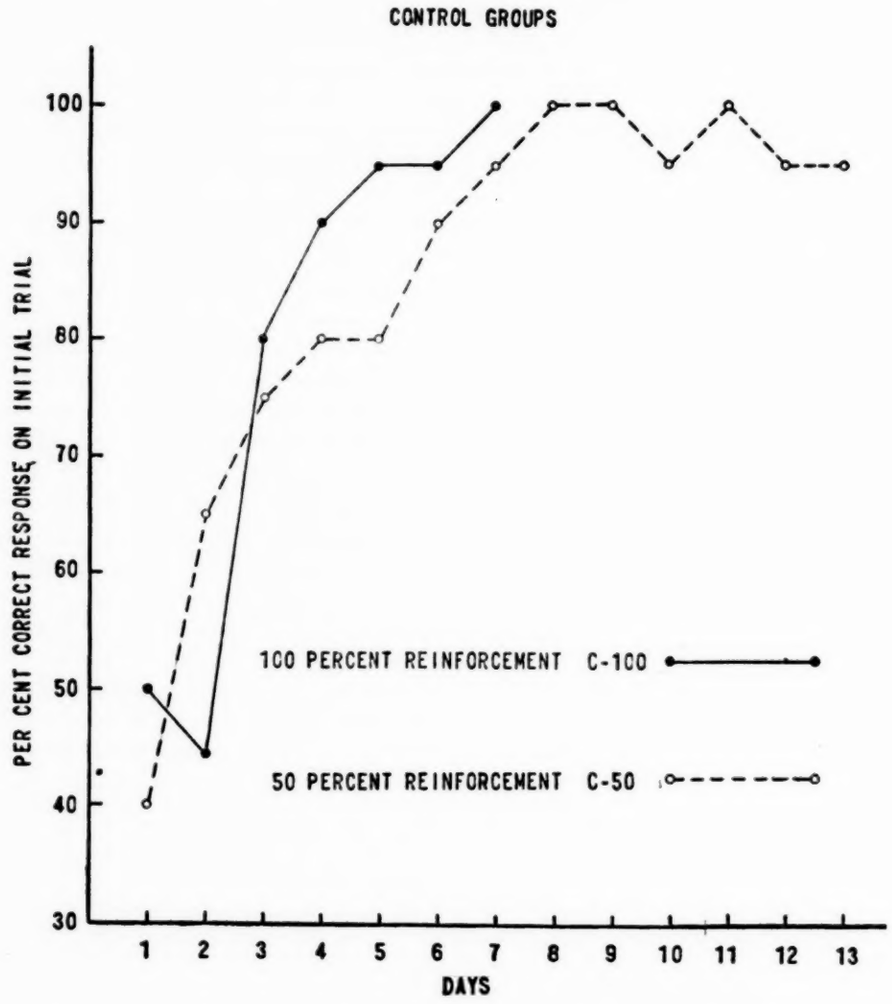


FIG. 2. Learning curves of the control groups, C–100 and C–50, as based upon the percent of correct responses on the initial trial of the day

an examination of the learning curves in the former it will be seen that, just as in previous experimental studies, the 100 percent reinforcement group (C–100) learned only slightly better than the partial reinforcement group (C–50). Table IV, which gives the mean number of correct responses made on the first trial of days 2–7,[6] shows that the slightly better overall learning performance of C–100 is not statistically significant. The *t* is only .52 which means that we

[6] The data of the first day are excluded because the variable of differential reinforcement did not operate until after the first day's run.

cannot reject the null hypothesis that there is no difference between the two groups.[7]

The results of this part of the experiment are seen, then, to be in essential agreement with the findings of Humphreys and others who report that partial reinforcement is just about as effective in producing learning as 100 percent reinforcement. It should be recalled that this finding is in accord with our theoretical expectations, since no

TABLE III

COMPARISON OF THE 50 AND 100 PERCENT GROUPS UNDER CONTROL CONDITIONS IN TERMS OF THE MEAN NUMBER OF CORRECT RESPONSES ON THE INITIAL TRIALS OF EACH DAY FOR DAYS 2 TO 7

| Group | *N* | No. Correct Responses | | Diff. | *t* | *P* |
|---|---|---|---|---|---|---|
| | | Mean | $\sigma$ | | | |
| C–100 | 20 | 5.05 | .98 | .20 | .52 | .60 |
| C–50 | 20 | 4.85 | 1.35 | | | |

TABLE IV

COMPARISON OF THE SUB-GROUPS OF C–100 AND X–100 ON EACH DAY OF LEARNING IN TERMS OF THE PERCENT OF CORRECT RESPONSES ON THE INITIAL TRIAL

| Day | C–100 | | X–100 | |
|---|---|---|---|---|
| | Sub-group 1 | Sub-group 2 | Sub-group 1 | Sub-group 2 |
| 1 | 60 | 40 | 58 | 42 |
| 2 | 30 | 60 | 50 | 67 |
| 3 | 80 | 80 | 83 | 83 |
| 4 | 90 | 90 | 83 | 58 |
| 5 | 90 | 100 | 58 | 67 |
| 6 | 100 | 90 | 75 | 92 |
| 7 | 100 | 100 | 92 | 92 |
| Mean Percent | 79 | 80 | 71 | 71 |

Note: Sub-group 2 received twice the number of trials to the wrong side per day as did Sub-group 1.

precautions were taken to preclude the operation of possible secondary reinforcement effects.

*C. Comparison of learning under the experimental condition.*—The learning curves for the two experimental groups, X–100 and X–50, are plotted in Fig. 3. As may be readily noted, they present a sharp

[7] Attention should be directed here to the fact that the data for the two sub-groups of the C–100 group, described in the section on experimental procedure, showed no differences in their learning scores and hence were combined. Not only are the means for the total learning period practically equal but the data are identical at almost all points along the learning curves (see Table IV).

contrast to the results obtained under the control condition. The 50 percent group exhibited a marked retardation in its rate of learning as compared with the 100 percent group (X–100). Whereas Group X–100 had attained a performance level of 92 percent by the seventh day, the *Ss* of Group C–50 were responding only 62.5 percent correctly.[8] Moreover, the latter group did not reach the 90 percent level of correct choice until the 12th day.

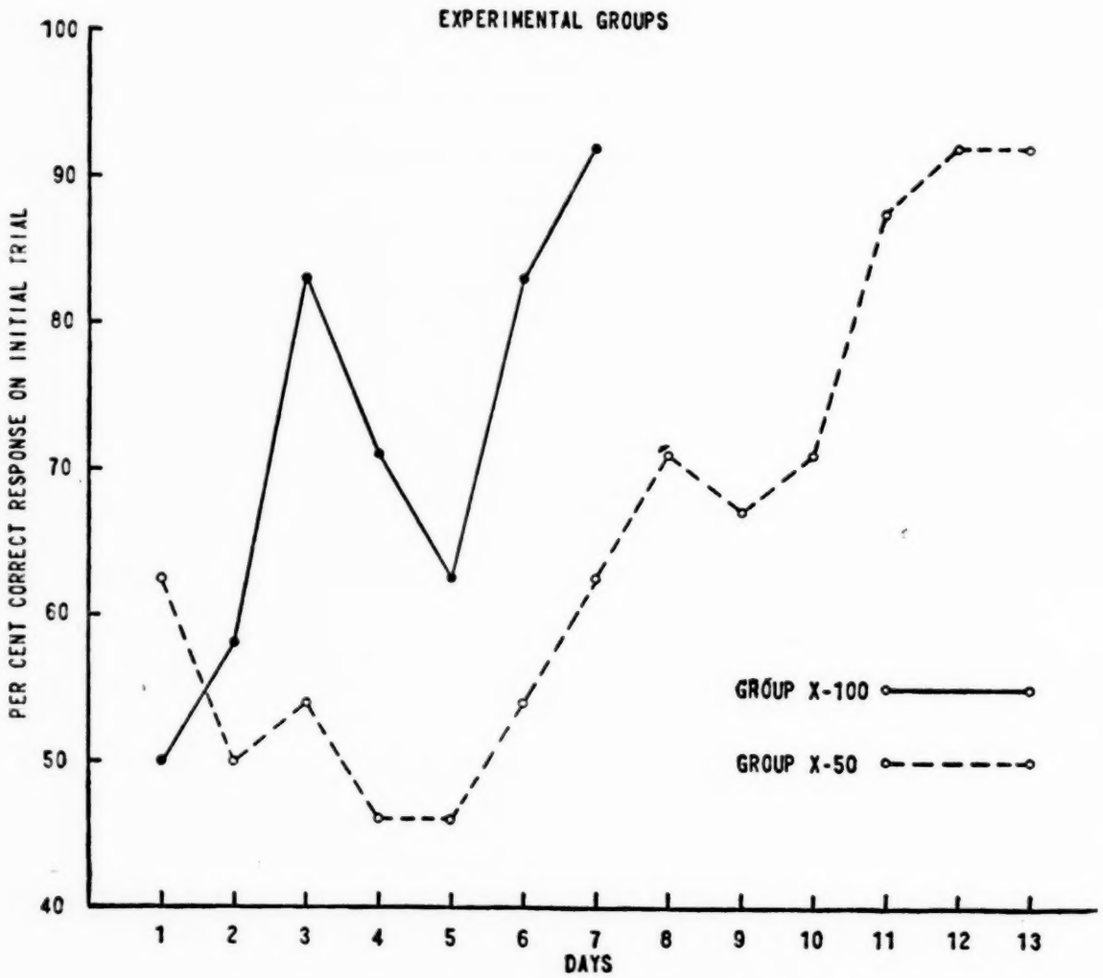


FIG. 3. Learning curves of the experimental groups, X–100 and X–50, as based upon the percent of correct responses on the initial trial of the day

It is also apparent from the data given in Table V that the overall learning scores for the two groups, based on days 2–7, are significantly different. The mean difference of 1.38 correct responses on the initial trials of this period gives a *t* of 3.06 which is significant beyond the one percent level of confidence.[9]

[8] This difference of 29.5 percent is significant at the two percent level as determined by means of the formula for the standard error of the difference between two uncorrelated percentages.

[9] Again the two sub-groups of the 100 percent reinforcement group showed no difference in their overall learning score or in their final level of performance and so were combined (see Table IV).

TABLE V

COMPARISON OF THE 50 AND 100 PERCENT GROUPS OF THE EXPERIMENTAL CONDITIONS IN TERMS OF THE MEAN NUMBER OF CORRECT RESPONSE ON THE INITIAL TRIALS OF EACH DAY FOR DAYS 2 TO 7

| Group | *N* | No. Correct Responses | | Diff. | *t* | *P* |
|---|---|---|---|---|---|---|
| | | Mean | $\sigma$ | | | |
| X–100 | 24 | 4.50 | 1.52 | 1.38 | 3.06 | <.01 |
| X–50 | 24 | 3.12 | 1.53 | | | |

That essentially the same results are obtained with the two free trial measure as with the first trial measure may be seen in Table VI. There is little difference between the 50 percent and 100 percent groups in the control condition. But in the experimental condition the difference between the partially and consistently reinforced groups is very marked.

TABLE VI

COMPARISON OF THE 50 PERCENT AND 100 PERCENT GROUPS OF BOTH THE CONTROL AND EXPERIMENTAL CONDITIONS FOR THE FIRST TWO FREE TRIALS ON DAYS 2–7 *

| Group | *N* | No. Correct Responses | | Diff. | *t* | *P* |
|---|---|---|---|---|---|---|
| | | Mean | $\sigma$ | | | |
| X–100 | 24 | 8.46 | 1.73 | 2.42 | 4.2 | <.01 |
| X–50 | 24 | 6.04 | 2.18 | | | |
| C–100 | 20 | 9.15 | 1.24 | .80 | 1.7 | .10 |
| C–50 | 20 | 8.35 | 1.49 | | | |

* A total of 11 responses.

*D. Comparison of learning in the experimental condition in terms of the number of previous food reinforcements.*—In Fig. 3 the learning curves are drawn in terms of the number of days. In other words, at each point along the abscissa the two groups of *Ss* have had the same amount of practice in making the correct response (two responses per day).

It is obvious from this graph that equal amounts of practice have not led to equality of learning in the two groups. If, however, we plot the performance of the two groups in terms of the number of previous reinforcements rather than the amount of practice, we find that the curves of the two groups are much more comparable. Fig. 4 represents such a plotting of the data. The two curves in this

graph represent the same data as in Fig. 3 except that the curve of Group X–100 is stretched out so as to equate the groups for number of previous reinforcements.

Examination of this figure reveals that the two groups attained approximately the same level of performance at the end of the training. Moreover, if we compare the total number of correct responses on the trials following 2, 4, 6, 8, 10, and 12 reinforcements we find no significant difference between the two groups.[10] That is, when the 50 percent and the 100 percent reinforcement groups are matched for number of reinforcements there is no difference either in the overall learning score or in the final level of performance.[11]

*E. Extinction data.*—Table VII presents the data for the 12 extinction trials in terms of the percentage of correct responses made during the series. Examination of the table reveals that there is no significant difference between the partial and consistently reinforced groups either under the control or experimental conditions.

TABLE VII

MEAN PERCENT OF FIRST RESPONSES CORRECT DURING THE EXTINCTION SERIES FOR BOTH THE EXPERIMENTAL AND CONTROL GROUPS

| Control | | Experimental | |
|---|---|---|---|
| C–50 | C–100 | X–50 | X–100 |
| 74.6 | 77.5 | 67.0 | 69.1 |

The failure to obtain a difference in the control condition requires further discussion because to a certain extent the finding is not in agreement with the results of Humphreys' experiment with the Skinner box. In his experiment, it will be recalled, there was a difference in favor of the partial reinforcement group in the extent to which the partial and consistently reinforced groups resisted extinction after having had an equal number of food reinforcements. The dis-

[10] The mean number of correct responses is 4.5 for the X–100 group and 4.1 for the X–50 group. The difference of .4 gives a $t$ of only 1.0 as compared with a value of 2.0 necessary to reject the null hypothesis at the five percent level of confidence.

[11] There is, however, an unusual feature about the learning curve of Group X–100, namely, the sharp rise on the third day and the large sharp drop on the fourth and fifth days. If the drop is assumed to be a systematic and not a chance phenomenon, as is indicated by the fact that the value of 62.5 percent on the fifth day would occur by chance once in a hundred times, it is probably to be accounted for in the following manner. In the act of switching the delay boxes and forcing the animals to equalize the number of responses to each side, both delay boxes gradually acquire an equal and potentially large amount of secondary reinforcement, and as the reinforcing power moves backward from the goal box to the delay box a reinforcing state of affairs now presumably occurs in the delay box, irrespective of what end box succeeds it. As a result when the animals are forced predominately to the wrong side on the last two or four trials, as of the third day, they are actually being reinforced on this wrong response. Consequently, with the tendency to go wrong stronger on the next day we find a smaller percentage of correct initial responses. Additional weight is given this systematic interpretation by the fact that the sub-group of X–100 which received a greater number of forced trials to the wrong side shows the drop sooner than the other sub-group (see Table IV).

crepancy is probably to be accounted for, in part, by the fact that a different type of extinction measure was employed, and in part by the fact that the two groups of Ss, C–100 and C–50, had about reached the limit of learning at the point at which they were extinguished, with the result that in the 50 percent group the extra trials involving secondary reinforcement could not produce an appreciable amount of further learning. In the context of Hull's theo-

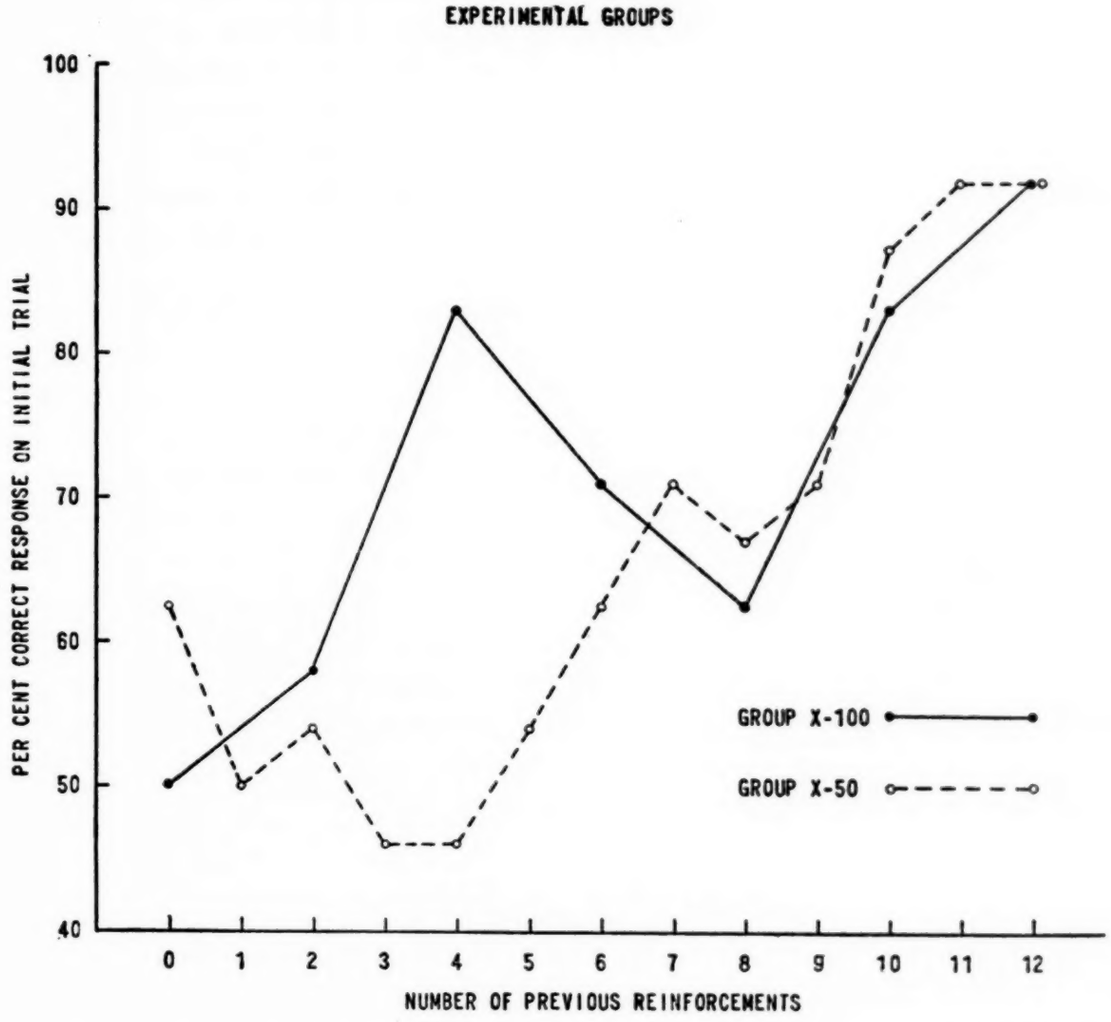


FIG. 4. Comparison of the initial trial learning curves of groups X–100 and X–50 when based upon an equal number of previous reinforcements

rizing about the limit of the curve of habit growth, this interpretation would require the assumption that the limit for the correct response habit was attained coincidentally with the attainment of 100 percent on the learning curve. Only in such circumstances would the additional secondary reinforcements received by the partially reinforced group not be effective in producing greater response strength.

A further finding revealed by the extinction data was the occurrence of a clear-cut difference between the animals extinguished with

the white goal box and those extinguished with the neutral gray end box. This result supports the finding of Bugelski (3) that resistance to extinction is greater when a sub-goal is present during the extinction series. The mean difference in the number of correct responses over the last half of the extinction series (trials 7–12) is 1.07 and the *t*, which is significant beyond the one percent level, is 3.02.

*F. Effect of extra forced responses to incorrect side.*—Mention has already been made of the absence of any difference between the learning of the sub-groups of C–100 and X–100, although one sub-group in each group received twice the number of non-reinforcements to the wrong side as did the other sub-group (see Table IV). This would tend to suggest that a differential number of non-reinforcements of the incorrect habit plays a relatively unimportant role in such simple trial and error learning.

This finding is in line with somewhat similar results reported by Spence (13). In the learning of a sequence of discrimination problems by chimpanzees, this investigator found no relation to exist between the number of differential non-reinforcements that the responses had received in the past and the subsequent response strength, *unless* the responses had previously been reinforced.

One possible factor offsetting the weakening effects of non-reinforcement in these extra trials in the present experiment is the fact that secondary reinforcing cues were present in the delay boxes. It is possible that this reinforcement compensated for the effects of the extra non-reinforcements. Further experimentation needs to be directed at this problem of the effects of non-reinforcement of responses in trial and error learning situations.

## Discussion of Results

The main experimental findings of the present investigation seem to be clear in their implications. The fact that a significant difference in learning was obtained between partial (50 percent) and consistent (100 percent) reinforcement groups when the experimental conditions were such as to minimize the effects of secondary reinforcement, whereas no difference was obtained between such groups under the more usual learning conditions, definitely suggests that secondary reinforcement was the responsible factor. That is, on trials in which no food reinforcement was given a secondary reward was nevertheless operative.

Actually, the present investigation is not the first to obtain a difference between the learning of partially and consistently reinforced groups. Using the straight alley maze and a latency measure,

Finger (5) found that his 100 percent groups showed significantly lower latencies at the completion of training than groups receiving reinforcement on 50 percent trials. These data are complicated, however, by the fact that the starting and food boxes were used interchangeably in these experiments. It is possible that this procedure reduced the secondary reward value of the boxes because of the fact that the animal had to spend a considerable amount of time in them in which food was not given; in fact, on half of the occasions the animal had to leave the box without receiving food. According to this conjecture, Finger's results are to be accounted for on the same basis as those obtained for our experimental condition, namely, the absence or weakness of secondary reward on the non-food trials.

Attention should be called here to certain as yet undiscussed aspects of Humphreys' experimental findings. With the exception of his eyelid conditioning study Humphreys has made use of resistance to extinction as his chief measure of the amount of learning. In the eyelid study he found, just as in the case of the control groups of the present study, no significant difference between the acquisition scores of the partially and consistently reinforced groups. His resistance to extinction measure, however, actually favored the partially reinforced group, a result which is somewhat out of line with his own conclusion that learning is a function of the number of practice trials.

This latter finding seems to be characteristic only of his classical conditioning studies (eyelid and galvanic skin response). It was not obtained in the Skinner-box experiment. Thus, reference to Table I will reveal that with the resistance to extinction measure there was no significant difference between the two groups (2 and 3) which received an equal number of practice trials and different proportions of reinforcements.[12]

Another difference that appears to be associated with this finding of greater resistance to extinction on the part of the partial reinforcement group after an equal number of training trials is that it has occurred in studies involving human subjects but not infra-human *Ss*.[13] It is possible that it is a function of some set which is predominantly verbal in nature. Humphreys has considered this possibility, but is of the opinion that verbal abilities are not the sole

[12] The difference referred to here should not be confused with the difference between groups 1 and 2 and groups 3 and 4 which is discussed above. The latter groups are compared when matched for number of reinforcements but not for number of practice trials.

[13] There is possibly some question concerning this difference between human and animal subjects so far as Skinner's experiments are concerned (12). Unfortunately none of his experiments provide a clear-cut answer to this question. Finger (5, 6) has reported the unusual finding that resistance to extinction in his set-up is inversely related to the number of reinforcements, although it does not seem to be a function of the consistency of reinforcement.

explanation of the result. The explanation of this phenomenon in terms of stimulus patterning proposed by Hull possibly plays a role.

In any event, it would appear from this investigation that Humphreys was somewhat over-hasty in drawing his conclusion that the strength of a learned response is a function of the number of practice trials. When proper care is taken to control possible secondary reinforcing cues it is found that the strength of the response is definitely a function of the number of reinforcements and not the number of practice trials.

## Summary and Conclusions

This study was designed to test the hypothesis that learning under partial reinforcement has been found to proceed in much the same manner as under 100 percent reinforcement because of the operation of secondary reinforcement on ostensibly non-reinforced trials. Learning with 50 percent reinforcement and 100 percent reinforcement was compared in control groups in which no attempt was made to control secondary reinforcement and in experimental groups in which conditions were arranged so as to minimize the effects of secondary reinforcement.

Control of secondary reinforcement was effected by the use of a new type of T-maze apparatus in which it was possible to control intra- and extra-maze cues. The 50 percent reinforcement groups were trained for 12 days, receiving one food reinforcement per day, while the 100 percent groups were trained for six days, receiving two food reinforcements per day. Eighty-eight animals, 40 control and 48 experimental, were employed.

The results in terms of overall learning scores (number of correct responses on test trials) and in terms of the final level of performance (percent correct responses) revealed no differences between the partial and continuous reinforcement groups in the control condition, whereas significant differences were found in the experimental condition. As a result of analysis of this and other related studies the following conclusions would appear to be warranted:

1. Learning is more accurately described as a function of the number of reinforcements than the number of practice trials.

2. Secondary reinforcement may operate in a partial reinforcement learning situation, serving to increase the number of effective reinforcements.

3. Resistance to extinction after either equal or unequal practice is not always greater with partial reinforcement than with consistent reinforcement.

4. Differential non-reinforcement of the incorrect response apparently has no effect on the rate of learning in a simple trial and error situation. This conclusion should be qualified, however, because of the small number of additional non-reinforcements given.

(Manuscript received December 5, 1945)

## References


1. Brogden, W. J. The effect of frequency of reinforcement upon the level of conditioning. *J. exp. Psychol.*, 1939, **24**, 419–431.
2. Brunswik, E. Probability as a determiner of rat behavior. *J. exp. Psychol.*, 1939, **25**, 175–197.
3. Bugelski, R. Extinction with and without sub-goal reinforcement. *J. comp. Psychol.*, 1939, **26**, 121–134.
4. Cole, L. E. A comparison of the factors of practice and knowledge of experimental procedure in conditioning the eyelid response of human subjects. *J. gen. Psychol.*, 1939, **20**, 349–373.
5. Finger, F. W. The effect of varying conditions of reinforcement upon a simple running response. *J. exp. Psychol.*, 1942, **30**, 53–68.
6. Finger, F. W. Retention and subsequent extinction of a simple running response following varying conditions of reinforcement. *J. exp. Psychol.*, 1942, **31**, 120–132.
7. Hull, C. L. *Psychology seminar memoranda*, 1940–1941. Bound mimeographed copies on file in the libraries of the Univ. of Iowa, Univ. of North Carolina, Yale Univ.
8. Hull, C. L. *Principles of behavior*. New York: D. Appleton-Century Co., 1943.
9. Humphreys, L. G. The effect of random alternation of reinforcement on the acquisition and extinction of conditioned eyelid reactions. *J. exp. Psychol.*, 1939, **25**, 141–158.
10. Humphreys, L. G. Extinction of conditioned psychogalvanic responses following two conditions of reinforcement. *J. exp. Psychol.*, 1940, **37**, 71–78.
11. Humphreys, L. G. The strength of a Thorndikian response as a function of the number of practice trials. *J. comp. Psychol.*, 1943, **35**, 101–110.
12. Skinner, B. F. *The behavior of organisms*. New York: D. Appleton-Century Co., 1938.
13. Spence, K. W. Analysis of the formation of visual discrimination habits in chimpanzees. *J. comp. Psychol.*, 1937, **23**, 77–100.

# DARK ADAPTATION AS A FACTOR IN THE SENSITIZATION OF THE BETA RESPONSE OF THE EYELID TO LIGHT[1]

BY DAVID A. GRANT AND EUGENIA B. NORRIS

*University of Wisconsin*

The present study was designed to investigate the influence of dark adaptation on the *sensitization* of the *β-response* of the eyelid to light. The β-response may be distinguished from the primary reflex to light, or α-response, by its longer latency and its greater susceptibility to sensitization. Sensitization is shown by reliable increases in frequency and amplitude of the response.

In previous studies (**2, 3**) eyelid responses to light which resembled conditioned eyelid responses were obtained under circumstances which precluded ordinary conditioning or pseudo-conditioning (**5**) reinforcement. These responses (β-responses) were studied systematically, and their frequency and amplitude were found to be influenced by the intensity of light stimulation (**4**). The frequency *increment* or degree of sensitization was found also to be influenced by stimulus intensity.

Although the previous studies have uniformly demonstrated sensitization of the β-response they were not designed to cast light on the nature of the sensitizing process. The present study was carried out to discover whether one of the factors common to the experiments in which sensitization occurred, namely, dark adaptation, could be shown to influence the course of sensitization of the β-response.

## Apparatus and Procedure

1. *Apparatus.*—The stimulating and recording apparatus was essentially the same as that used in an earlier study (**4**). It consisted principally of a modified Dodge photochronograph (1) which recorded movements of the right eyelid by means of a simple lever system.

The *S* rested his head against a padded head-rest and gripped a bite-board lightly with his teeth. He looked straight ahead into a box-like enclosure which was painted flat black inside and out. Only one stimulus was used. It consisted of the illumination to a brightness of 241 millilamberts of a circular milk-glass window, 10 cm. in diameter, located at the back of the enclosure, 60 cm. directly in front of the *S*s eyes. This value of the stimulus was selected to maximize sensitization of the β-response (**4**). The duration of the stimulus was about 750 millisec. The circular stimulus window was in the center of a 30 cm. square sheet of light-colored drawing paper which formed a light-diffusing mat surface. For dark adaptation the enclosure remained

[1] This research was supported in part by a grant-in-aid from the Graduate School of the University of Wisconsin for the year 1945–46.

dark except for a very dim (.01 millilamberts) pre-stimulus illumination on the milk-glass. The interior of the enclosure could be illuminated so that the light-diffusing square had a brightness, varying from place to place on its surface, from three to six millilamberts. Under these circumstances, the milk-glass window itself was illuminated to a brightness of 2.4 millilamberts.

2. *Experimental design.*—Two procedures were used. They were a *dark adaptation* procedure in which the interior of the enclosure was kept dark except for the pre-stimulus brightness of the milk-glass window, and a *non-adaptation* or *control* procedure, in which the enclosure was illuminated except during stimulation and recording. The procedures may be summarized as follows:

| Dark Adaptation | Non-adaptation or Control |
|---|---|
| Test Series 1<br>(Enclosure dark—4 light stimuli and one control trial) | Test Series 1<br>(Same as Dark Adaptation procedure) |
| Fixation (7 min.)<br>(Enclosure dark. *S* looks at fixation spot on milk-glass window) | Fixation (7 min.)<br>(Enclosure illuminated first six min.—dark last min. *S* looks at fixation spot on milk-glass) |
| Test Series 2<br>(Same as Test Series 1) | Test Series 2<br>(Same as Test Series 1) |
| Rest (2 min.)<br>(Enclosure dark) | Rest (2 min.)<br>(Enclosure illuminated) |
| Fixation (7 min.)<br>(Same as above) | Fixation (7 min.)<br>(Same as above) |
| Test Series 3<br>(Same as Test Series 1) | Test Series 3<br>(Same as Test Series 1) |
| Rest (2 min.)<br>(Same as above) | Rest (2 min.)<br>(Same as above) |
| Fixation (7 min.)<br>(Same as above) | Fixation (7 min.)<br>(Same as above) |
| Test Series 4<br>(Same as Test Series 1) | Test Series 4<br>(Same as Test Series 1) |

In each of the above *Test Series* two light stimuli, one control trial, and two more light stimuli were given in that order. The control trial consisted of operating all apparatus and recording eyelid movements as on stimulus trials, except that the light stimulus was not presented. This served as a check on the presence of incidental conditioning. The interval between trials was varied randomly between 25 and 45 sec., and no 'ready' signal was given.

3. *Subjects.*—The *S*s were men and women students in the elementary course in psychology during the summer of 1945 who were asked to volunteer. *S*s were assigned to the two groups alternately according to their order of appearing for the experiments. There were 33 *S*s in the Dark Adaptation group and 32 *S*s in the Non-adaptation group.

## Results

The latencies of all eyelid responses recorded in the present study formed the bimodal distribution found to be characteristic of the eyelid responses evoked by moderately bright light stimuli (2, 3, 4). Following the definition given in an earlier report (4, p. 395), the reflex responses which ranged in latency from 40 to 110 msec. were

called $\alpha$-responses, and the responses which ranged from 120 to 240 msec. were called $\beta$-responses. (As pointed out in earlier papers, the $\beta$-responses lie within the latency range of conditioned eyelid responses.)

1. *Effect of dark adaptation on frequency of $\beta$-responses.*—The percent frequency of $\beta$-responses occurring on each Test Series on the experiment is given in Table I for the Dark Adaptation and the Non-adaptation groups. Each percentage is based upon four trials per *S* for each Test Series, making a total of 136 trials for each entry in the Dark Adaptation group and 128 trials for each entry in the Non-adaptation group.[2]

TABLE I

PERCENTAGE FREQUENCY OF BETA RESPONSES DURING THE FOUR TEST SERIES OF THE EXPERIMENT

| Group | Test Series | | | |
|---|---|---|---|---|
| | 1 | 2 | 3 | 4 |
| Dark adaptation..... | 13.2 | 30.0 | 45.6 | 51.5 |
| Non-adaptation..... | 9.4 | 20.3 | 11.7 | 22.7 |
| Percent confidence level of difference..... | <30.0% | <0.1%* | <0.1%* | <0.1%* |

* Difference on Test Series 1 held constant by analysis of covariance.

The data show little difference between groups at the beginning of the experiment (Test Series 1). In the later Series the percent frequency of $\beta$-responses becomes progressively greater in the Dark Adaptation group, whereas there is little or no upward trend in the Non-adaptation group.

The initial difference between groups was not statistically significant ($t = 1.10$). This difference, however, favored the Dark Adaptation group, so that it was thought advisable to eliminate its influence in the statistical analysis of later differences. This was accomplished by means of the analysis of covariance (**8**). The percent confidence level for each difference is given in the bottom line of Table I. The analyses showed that differences on Test Series 2, 3, and 4 were highly significant ($P < 0.1\%$). This indicates clearly that dark adapted *S*s give a higher frequency of $\beta$-responses than non-adapted *S*s when both are tested under otherwise similar conditions.

The over-all *increase* in frequency from Test Series 1 to Test Series 4 was, of course, greater in the Dark Adaptation group. The

[2] In computing sampling statistics, data were re-grouped according to *S*, so that the sampling *N* for the Dark Adaptation group was 33, and the *N* for the Non-adaptation group was 32.

difference in increments was investigated by means of the *t*-test and was found to be statistically significant at the one percent level ($t = 2.84$). Thus dark adaptation was shown to produce *greater increment* in frequency of $\beta$-responses than maintenance of fixation without adaptation.

It is of interest to note that the greatest frequency increment in the Dark Adaptation group appears from Test Series 1 to Test Series 2 ($t = 4.14$). The increments from Test Series 2 to 3 and from Test Series 3 to 4 were not significant ($t = 1.34$ and 1.14, respectively).

2. *Effect of dark adaptation on magnitude of $\beta$-responses.*—The magnitude [3] of $\beta$-responses on each of the Test Series of the experiment is given in Table II. These data follow the same trends as the

TABLE II

AVERAGE MAGNITUDE (MM.) OF BETA RESPONSES DURING THE FOUR TEST SERIES OF THE EXPERIMENT

| Group | Test Series | | | |
|---|---|---|---|---|
| | 1 | 2 | 3 | 4 |
| Dark adaptation..... | 1.31 | 5.31 | 5.45 | 5.62 |
| Non-adaptation..... | 1.28 | 1.79 | 1.56 | 2.56 |
| Percent confidence level of difference..... | 50.0% | <1.0%* | <0.1%* | <2.5%* |

* Difference on Test Series 1 held constant by analysis of covariance.

frequency data of Table I. Except for Series 1, the differences between the Dark Adaptation and Non-adaptation groups were statistically significant as shown by the percent confidence levels from the *t*-test on Test Series 1 and the analyses of covariance on Series 2, 3, and 4. These confidence levels are given in the bottom line of Table II. A *t*-test of the difference in over-all gain in magnitude from Test Series 1 to Test Series 4 showed that dark adaptation produced significantly greater ($t = 2.11$; $P < 5\%$) gain in magnitude than did visual fixation without adaptation.

3. *Effect of dark adaptation on the frequency and magnitude of the $\alpha$-response.*—The frequency and magnitude of primary reflexes to light or $\alpha$-responses showed typical reflex adaptation drops (6) from Test Series 1 to Test Series 4. In frequency, magnitude, and general trends the $\alpha$-responses of the Dark Adaptation group were similar to those of the Non-adapted group. There was a slightly

[3] Magnitude (7) is obtained by averaging amplitude (in mm.) for each *S* for each Test Series. The average for each *S* is then used as his score to compute the group average magnitude. Absence of response is counted as zero magnitude, so that magnitude is positively correlated with frequency.

greater drop in frequency for the Non-adaptation group, but the difference between groups was not significant ($t = 1.04$). The fact that there was no difference in $\alpha$-response trends, in spite of a significant difference in $\beta$-response trends, shows the functional difference between the $\alpha$-response and the $\beta$-response of the eyelid to light. This difference confirms in all respects the earlier observations to this effect (4).

## Discussion

1. *The process of sensitization in the β-response.*—The results of the present study show clearly that dark adaptation is an important factor in the sensitization of the $\beta$-response of the eyelid to light. The sensitizing effect of dark adaptation fits in, furthermore, with the findings of previous investigations. In each of the previous studies showing sensitization of the secondary eyelid responses (2, 3, 4) conditions favored dark adaptation during the course of the experiment. In addition, the fact that frequency of $\beta$-responses is greater when the stimulus is more intense (4) would suggest that dark adaptation with the consequent increase in effective strength of the stimulus should result in frequency increments.

The increase in effective visual sensitivity follows an essentially logarithmic course during dark adaptation. If the sensitization of the $\beta$-response is brought about chiefly by dark adaptation, then the frequency of $\beta$-responses might be expected to follow a logarithmic curve. Furthermore, since the previous study has shown the frequency of $\beta$-responses to be linearly related to the logarithm of *stimulus intensity*, this also could be included in the logarithmic equation describing sensitization. The equation might therefore be of the form $F = a \log It + b$, where $F$ is percent frequency, $I$ is intensity (measured, at present, in millilamberts), $t$ is time in sec., and $a$ and $b$ are constants.

In Fig. 1 the logarithmic function has been fitted to the frequency data of the Dark Adapted group. The black dots represent the experimental results. The chi-square test of goodness of fit (8) gave $\chi^2 = 1.626$, which, for two degrees of freedom, shows that the observed data did not depart significantly from the curve.

The equation with its present constants fits surprisingly well the 582 millilambert curve of the previous study (4, p. 397), but it does not fit the curves for the lesser intensities. A more general function such as $F = \log I^a t^b + c$ may be required, and a complete formulation for dark adaptation would have to include such variables as pre-adaptation and adaptation illumination.

Factors other than dark adaptation may participate in the

mechanism of sensitization of the β-response. This is suggested by the fact that the increase in percent frequency of β-responses in the Non-adaptation group from Test Series 1 to Test Series 4 was significant at the five percent confidence level ($t = 2.64$). Too much confidence should not be placed in this marginally significant trend, however, because of the inversion from Test Series 2 to Test Series 3, where the frequency decreases (*cf.* Table I). The control group results emphasize the fact that further work is needed before it may

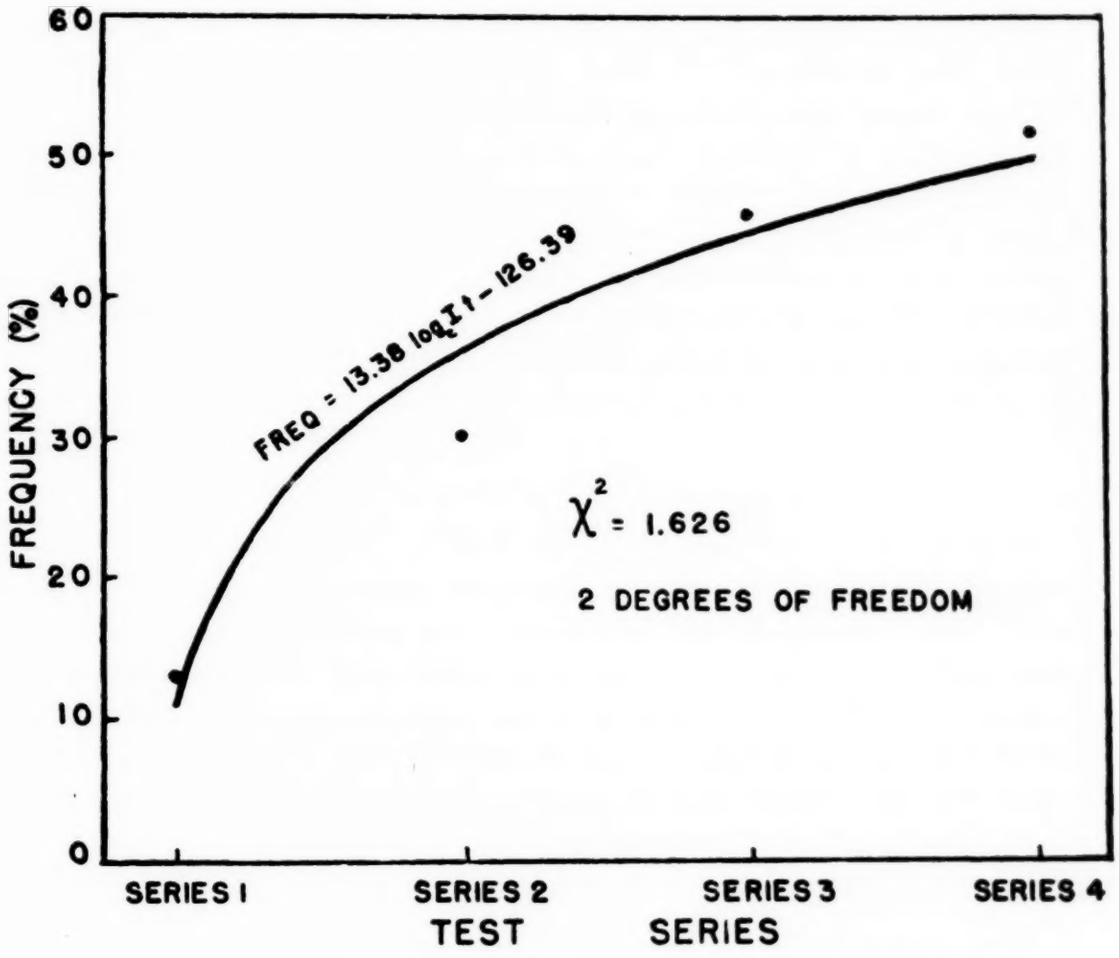


FIG. 1. Frequency of β-responses in the Dark Adaptation group with fitted logarithmic curve

be said that the β-response is sensitized by virtue of dark adaptation alone.

2. *Implications of findings for the conditioned eyelid response.*—Due largely to the excellence and convenience of the technique developed by Dodge and Hilgard and Marquis, many studies of the conditioned eyelid response have been carried out using light as the conditioned stimulus and using the photochronograph as the recording instrument. This procedure, with few exceptions (9), has incidentally involved dark adaptation on the part of the *S*s, and it therefore seems

likely that the secondary response of the eyelid to light ($\beta$-response) has entered into the findings of these studies.

There are at least two plausible ways in which the $\beta$-response may have affected earlier CR results. First, $\beta$-responses may easily be confused with CR's (3) so that some findings reported for the eyelid CR may not really be associative or learned phenomena. Such confusion of $\beta$-responses with CR's would presumably be eliminated, or at least greatly reduced, if pseudo-conditioning or visual fixation controls were run, or if conditioned discrimination were demonstrated. In addition to these techniques, low intensity of the conditioned stimulus and avoidance of dark adaptation would reduce the proportion of $\beta$-response sensitization in the data.

Second, the $\beta$-response may play a significant role in the establishment of the eyelid CR, so that dark adaptation would greatly facilitate conditioning.[4] These two alternatives are not mutually exclusive, and further research is needed to demonstrate the precise relationship of the $\beta$-response to the CR. The demonstration that dark adaptation and stimulus intensity are related to the sensitization of the $\beta$-response provides a foundation stone for this work.

## Summary

The purpose of the present study was to investigate the influence of dark adaptation on the sensitization of the $\beta$-response of the eyelid to light. Two groups of *S*s were used; one group ($N = 33$) was *dark adapted* and the other ($N = 32$) was subjected to a *non-adaptation* procedure. Eyelid responses of both groups were recorded under identical conditions at four times during the experiment. The light stimulus was presented and responses were recorded by means of a modified Dodge photochronograph.

The results of the experiment may be summarized as follows:

1. The latencies of the eyelid responses formed the bimodal distribution which was reported in earlier studies. The first mode consisted of primary reflexes or $\alpha$-responses to light, and the second mode consisted of $\beta$-responses which ranged in latency from 120 to 240 msec.

2. The $\beta$-responses of the Dark Adaptation group showed significantly greater frequency and magnitude than those of the Non-adapted group.

3. There was significantly greater sensitization (frequency and magnitude increments) in the $\beta$-responses of the Dark Adaptation group.

[4] Research on this problem is now in progress in the University of Wisconsin laboratory.

4. The trends of the $\beta$-responses were not related to those of the $\alpha$-responses. The $\alpha$-responses showed typical reflex adaptation.

It was thus shown that dark adaptation plays an important role in the sensitization of the $\beta$-response, and it was pointed out that this factor has serious implications for past and future work on the conditioned eyelid response where light is used as the conditioned stimulus.

(Manuscript received December 10, 1945)

## References

1. Dodge, R. A pendulum photochronograph. *J. exp. Psychol.*, 1926, **9**, 155–161.
2. Grant, D. A. The pseudo-conditioned eyelid response. *J. exp. Psychol.*, 1943, **32**, 139–149.
3. Grant, D. A. Sensitization and association in eyelid conditioning. *J. exp. Psychol.*, 1943, **32**, 201–212.
4. Grant, D. A. A sensitized eyelid reaction related to the conditioned eyelid response. *J. exp. Psychol.*, 1945, **35**, 393–402.
5. Grether, W. F. Pseudo-conditioning without paired stimulation encountered in attempted backward conditioning. *J. comp. Psychol.*, 1938, **25**, 91–96.
6. Hilgard, E. R., & Campbell, A. A. The course of acquisition and retention of conditioned eyelid responses in man. *J. exp. Psychol.*, 1936, **19**, 227–247.
7. Humphreys, L. G. Measures of strength of conditioned eyelid responses. *J. gen. Psychol.*, 1943, **29**, 101–111.
8. Lindquist, E. F. *Statistical analysis in educational research.* Boston: Houghton-Mifflin, 1940.
9. Weber, H., & Wendt, G. R. Conditioning of eyelid closure with various conditions of reinforcement. *J. exp. Psychol.*, 1942, **30**, 114–124.

# LEVEL OF ASPIRATION: AMBITION OR DEFENSE?

BY ROBERT R. HOLT

*Winter General Hospital, Topeka, Kansas*

## Statement of the Problem

Among experimental techniques for studying personality, level of aspiration is both one of the earliest and most popular. It is not surprising that it was eagerly taken up, for in a field where much was vague, it was precise; where quantification of results was difficult or impossible, it gave rise naturally to numerical scores; where subjectivity of scoring was the rule, it was objective and operational. And from the first, investigators have been sure that they had a measure of something important. The very name that Frank (**6**) gave to the principal variable in his modification of Hoppe's original technique (**14**) led naturally to the assumption that it gauges aspiration, ambition, or some similar trait. One of Hoppe's conclusions had been that the striking individual differences in *Anspruchsniveau* were probably related to such traits as ambition, prudence, and self-confidence. Jucknat (**16**) and Frank added other hypotheses to the list, and the latter investigator introduced the idea that level of aspiration itself was a generalized personality trait (**6**). 'Level of aspiration,' used in the sense of *goal* or *motivation*, has become one of the clichés of psychologists' conversations; May and Doob have even dignified the assumption involved in this usage by formulating a law, that all motivation is a function of the discrepancy between the level of achievement and the level of aspiration (**18**).

Later workers with level of aspiration have not been so sure that its significance was primarily as a diagnostic sign for ambition, or a 'level of goal-striving.' Gould early suggested that levels of aspiration might better be regarded as aspects of the person's typical modes of ego-defense:[1]

> It would seem doubtful . . . whether . . . one obtains an accurate measure of the individual's level of strivings. Rather,

[1] A preferable expression is *esteem-defense*. Since defense is one of the groups of functions that comprise the ego, the term 'ego-defense' is redundant. Furthermore, 'esteem-defense' makes it explicit that interest is focussed on those defensive processes which protect the person from loss of self-esteem and recognition from others (prestige, etc.) as distinguished from defenses against 'instinctual anxieties' or those concerned with loss of bodily integrity (see Freud, **8**).

> the average difference score [2] would appear to be more nearly representative of a kind of protective mechanism against failure, than indicative of differences in the height of ['real'] aspiration level (11, p. 83).

Other authors too, such as Gardner (10), Rotter (21), and Sears (22), have found it necessary to interpret level of aspiration behavior in terms of processes having to do with defense of esteem.

How may one decide between these two conflicting interpretations? First of all, it should be recognized that both may be valid, since such a congeries of different behaviors are often included under the general heading, 'level of aspiration,' and since any one may be overdetermined. That is, it might be possible to devise experimental conditions in which statements of a certain kind about future performance would be more a function of the *S*s' degree of achievant motivation than of anything else. Correspondingly, other statements in other situations—different, but not so different that they could not still be called levels of aspiration—might be relatively independent of the need for achievement, and more than anything else expressions of esteem-defenses.

For any particular experiment, however, it should be possible to reject or accept either the 'incentive-hypothesis' (that levels of aspiration are primarily measures of striving) or the 'defense-hypothesis' (that they are primarily means of esteem-defense). Two methods of procedure suggest themselves. First: to deduce, for any particular experimental set-up, the consequences which would follow on either hypothesis, and to decide on the basis of the experimental finding. Second: to see whether the levels of aspiration obtained correlate with generalized characteristics of personality that have to do mainly with goal-seeking (for example, needs for achievement or construction) or defense of esteem (for example, pride, or need for defendance). The present paper deals with the former method, but it may be stated that the results of the second method in a different experimental set-up favor the defense-hypothesis.[3]

*Deduction from incentive-hypothesis.*—Assume that the height of the stated level of aspiration is a simple function of the intensity of motivation. Two *S*s, H and L, are engaged in a task. They have equal ability to do what the task requires, but when asked about their future performance, H expresses a higher level of aspiration than

[2] The difference score, or D-score, is the discrepancy between an expressed aspiration and a preceding level of performance which presumably forms an important part of the frame of reference in which the estimate of future performance is formed. It is the most widely used measure of level of aspiration.

[3] For a discussion of the evidence from personality correlates, see Holt (12). Some of the relevant results have been summarized in another paper (13).

L. The motivation of H therefore must be greater: the greater magnitude of the force propelling him towards his goal has caused him to set a higher goal or expectation. Assume still further a simple relation between motivation and actual effort expended in work; then H should exert himself more than L in the activity that both undertake. Since both have equivalent ability, greater exercise of that ability by H must result in his achieving more. It follows, therefore, that when ability is held constant, there should be a positive correlation between level of aspiration and subsequent achievement.

*Deduction from defense-hypothesis.*—Assume that the level of aspiration is a form of verbal behavior by which the person can affect the social definition of the situation, as he is about to perform a task. Let it be granted further that his performance, its success or failure, will affect to some extent his opinion of himself and his standing in the eyes of others. Now, a level of aspiration of any size may be part of a pattern of esteem-defense. If the person relies on overt modesty to win approval, for example, he may give a low level of aspiration. If he hopes to be admired for expecting much of himself, he may give a high aspiration. So, unless it is assumed that one of these general sorts of defenses is used to a greater extent by most persons when they are highly motivated (an assumption for which there is no available evidence) the height of the level of aspiration should be independent of the strength of motivation on the part of the person giving it. Assuming as before that effort expended, and therefore achievement, will be highly correlated with degree of motivation, there is no reason to expect anything more than a chance relationship between level of aspiration and subsequent achievement, when ability is held constant. Knowing just what we do about H and L, that they are equally able and that the latter has expressed the lower level of aspiration, one would be unable to make any predictions about their relative achievement in the task both are to undertake. Virtually the only positive prediction that can be made on the defense-hypothesis is that there will be an association between uniformities of aspiration-giving and of other kinds of esteem-defensive behavior.

By obtaining levels of aspiration from *S*s before they undertake a task and correlating them with actual attainment in that task, ability being held constant, one can make a test of the two hypotheses about the meaning of levels of aspiration—a test that is not altogether crucial, but one that should furnish evidence for a choice.

*Previous experimentation.*—There have been a few reports of the correlation between statements about future performances and subsequent achievement, though only one experiment has been done in such a way as to make possible the kind of test proposed here.

The experiments of Filter (**5**) and Kneeland (**17**), in which slight relations between estimates and actual performance were obtained, did not provide for control of ability or previous achievement.

In the investigations reported by Bayton, on the other hand, control of ability was carefully carried out by the method of matched pairs (**3**). *Ss* were introduced to two different tasks, cancellation and arithmetic problems, and were asked which they would prefer to excel at. All chose arithmetic, which was therefore considered to be more ego-involved. After the first seven of the 11 trials in each task, the *E* reported a standard series of fictitious scores. Actual times were recorded, and *Ss* were matched for ability on the basis of their performances on trials six, seven and eight. After the eighth trial, *Ss* were not told their scores, but were asked to estimate them. Before comparing subsequent actual performances of *Ss* having high and low aspirations, the *E* matched them for these estimates of past performances as well as for actual achievement. Three aspirations were asked for after each of the last three trials, following estimates of past performance: *Maximum:* "the best score you think you could ever make in this task"; *Least:* "that score below which you are certain you will not fall if you should have a bad trial"; and *Actual:* "what you feel you will do this next time, exactly what you expect to make."

A trial-by-trial analysis of the two tests showed no reliable differences in actual performances on the trial immediately following the aspirations for the matched groups having high and low Maximum or Least estimates. With respect to the Actual estimates, only in the more-involved arithmetic test were there sizeable differences in performance between the groups giving high and low aspirations. *Ss* who gave high Actual estimates did better on the following trials than those who did not expect to do so well, but the differences were not reliable for any one trial. When all three trials were combined, the probability that such large differences would be obtained by random sampling of a uniform population was between .02 and .05. This finding was therefore adjudged significant.

In discussing the results of his experiment, Bayton concluded that his hypothesis was sustained: that the needs which cause high aspirations persist and cause better performance. He admits that the differences in his data are not large, and suggests three possible explanations. *First*, "the arithmetic task did not create enough ego-involvement to create a striking effect." *Second*, "these explicit statements of aspiration are not reliable measures of subjective goal-striving" but might instead be "either expressions of desire to do well or of merely wanting to appear ambitious." This seems clearly to be a recognition of the possible esteem-defensive function of levels of

aspiration. *Third*, the method of asking for estimates might be at fault. "Perhaps, as Irwin suggests, a key to future experimentation would be the shift from asking for *expectations* concerning future performances to asking the *S*s to state what they are *trying* to achieve" (3, p. 17).

Now, it happens that although the investigations reported here were planned and executed before Bayton's work had appeared, their design takes into account all three of the factors mentioned by him. On the first point, ego-involvement, the tasks used may be assumed to be considerably more likely to arouse the *S*s' anticipations of gain or loss of self-esteem.[4] Instead of administering miniature tests in a laboratory setting, I decided to take advantage of an accessible testing situation of real life: the college examination.

Indeed, a test in a college course offers a real opportunity for research on level of aspiration. Here is a situation of considerable importance to most students, about which they spend much of their time thinking and worrying and planning, and which is very widely taken as a measure of personal worth. Examination grades are the chief proofs that a son or daughter can offer to his or her parents that progress is being made academically, and that therefore love and money should be forthcoming. Furthermore, it is a natural thing to ask a student, "How are you going to do on the exam?"—much closer to his everyday experience than to get him to predict, trial by trial, his performance on a manual or quasi-intellectual task, a procedure that some *S*s find quite silly. Finally, both performance and aspiration are usually defined in terms of a metric scale, to a much greater extent than is true of other goal-seeking behavior in real life.

Bayton's second point, in slightly different guise, was made the crux of the present experiment. In addition, an attempt was made with one group of *S*s to get other data on defensive processes involved in making the estimates.

Finally, the matter of instructions. From the beginning there has been a great deal of variety in the experimental question used to elicit aspirations. Whether the *S* was asked: "What will you do?" (Gould, 11); "Where do you expect to stand?" (Gardner, 9); "What are you going to try for?" (Sears, 22), or what not, the answers were called levels of aspiration. The first explicit recognition of this heterogeneity was in a footnote in Gould's monograph.[5] She was quite aware that different *S*s interpreted her instructions differently,

[4] The definition of ego-involvement and its intimate relation to self-esteem are discussed in Holt (13).

[5] Later studies by Festinger (4) and Irwin & Mintzer (15) have dealt with this problem, with findings similar to mine.

some taking them as a call for an expectation, others for a hope, and so on, but she did not think such differences important, because:

> Current work with about 30 subjects indicates no difference whatever between response to "What will you do next time?" and "What do you intend to do next time?" (11, p. 42).

Nevertheless, it seemed to me that her test had not been conclusive, and that there was reason to expect that different kinds of levels of aspiration might have different functional significances. Therefore, in the first group of *S*s in the present experiment, half were asked for *expectations*, the other half for *goals*. Data from this group thus afford a direct test of Irwin's suggestion (as reported by Bayton, **3**, p. 3) that levels of aspiration in terms of what *S*s were *trying* to make would show more clearly the incentive value of the level of aspiration, if any.

From the standpoint of the tasks chosen and the *S*s' involvement in them, and of the instructions used, it is reasonable to expect a more significant result than that found by Bayton, but in the same direction—according to the incentive-hypothesis.

## Subjects

There were two groups of *S*s: students in the Harvard and Radcliffe course in Introductory Psychology, and students in the Harvard undergraduate course, Physics C. The first group comprised 19 Harvard Students in two class sections, and 57 Radcliffe students in four sections. Of this number, six were discarded, since complete information was not available for them, leaving 70 (18 men, 52 women).

The second group consisted of 109 Harvard undergraduates, of whom four were discarded because of evidence of not taking the experiment seriously, and seven because all needed data were not available, leaving 98 *S*s. The Physics C students formed a particularly good group for the study. The course is the more advanced of two introductions to physics offered by the department, and is typically the first college course in physics for undergraduates who are going to medical school, or who intend to concentrate in one of the natural sciences, since it is required for concentration in physics and several other departments. All classes were represented in the course, but most members were freshmen and sophomores; for all who take it, it is likely to be one of the most difficult and important courses that they are attending at the time. The kind of ego-involvement that is produced by these facts is evident from written

comments that were gathered from many of the Ss. One of the most articulate put it well:

> It is necessary to obtain at least a C in Physics C for two reasons, 1) it is a requirement for Medical School, 2) it is a requirement for a major in Biology. I assume that the higher the grade you will get, the better off you will be, as examinations are the only way they have, other than aptitude tests, and personal contact, to rate you with others applying for the same position or heading for the same goal. I would consider the mark 100 as possible, but improbable. ∴ I place down 90.

## Procedure

Data were obtained in slightly different ways from students in the two courses.

*Groups E and G (Psychology).*—In the section meeting during the week preceding the spring hour examination (March 1941), the *E* was introduced by the instructor of each section as a member of psychology department who was conducting an investigation. *E* then said that he was interested in students' attitudes towards grades, and would like their serious cooperation. Blank slips of paper were passed out; the Ss in any one section were then given either Instruction E or Instruction G. They were asked to sign their names, and were assured that the information would be confidential, that no member of the faculty would see any of the papers. An equal number of Harvard and Radcliffe sections were given Instructions E and G.

Instruction E, though spoken extemporaneously, was substantially as follows:

> Please write on the paper the letter grade that you *expect* to get on the coming hour exam; that is, the grade you think you *actually will* get. Please do not let anyone else see your paper, but make this your own honest estimate.

If there were any questions, the instructions were restated in similar words and explained until clear. Ss who received this instruction are referred to as *Group E.*

Instruction G was given in the same way, differing only in the first sentence, thus:

> Please write on the paper the letter grade that you are *trying* to make on the coming hour exam; that is, the grade you want to make, and that is your *goal.* Please do not let anyone . . . (etc.).

A few Ss asked if the grade they expected was the one they wanted;

they were told no, and the instructions were repeated in paraphrase. The sections to which this instruction was given make up *Group G.*

The *E* obtained the record of each *S* on the preceding hour examinations from the instructor, as well as the grade which each *S* made upon the hour examination for which he had made the prediction. Since the grades which the *S*s had received previously had been in terms of the A to E lettering system, estimates were asked for in these terms. For quantitative treatment, however, all grades were converted into numerical scores according to the following system. A plus was equated to 100, A to 95, A minus to 90, B plus to 85, and so on down to E, equated to 35. These numbers very closely approximated the actual equivalent grade points on the hour examinations. The translation used here assumes that the intervals between letter grades, and between the plus-minus subdivisions, are perceived as equal. The legitimacy of this assumption is not proved, but is at least made credible, by the close correspondence of results from all groups, and by the agreement of results when Chi-square and the analysis of variance are used (see *Results*, below).

*Group C* (*Physics*).—When the students came into the lecture hall the week before the final examination in the course, each man found on his seat a slip of paper reading as follows:

> Please write below the numerical grade (from 0 to 100) that you are trying to make on the final examination in Physics C. This information is strictly confidential, no one connected with the Physics faculty will see it. Please do not confer with anyone else on this, but make it *your own* goal, that *you* are trying for.
> The grade you are trying to make on the final.....
> Your name.....

The lecturer cooperated in using his prestige to call the attention of the course to the slips of paper before beginning the lecture, saying that it was not a joke but part of an investigation being carried out by the psychology department on students' attitudes.[6] He urged them to fill out the slip quickly and accurately, putting down their classes as well as their names, with any comments they felt like making, and stressed again the serious nature of the investigation. Thereupon he went on with the lecture. Papers were collected at the end of the lecture by the *E*. These *S*s will be referred to as *Group C.*

After the examination, grades for the *S*s on all examinations throughout the year, including the one for which they had predicted,

[6] I would like to express my thanks to Professor F. A. Saunders for his kind and willing assistance, which made this experiment possible; likewise to Professor G. W. Allport, who kindly gave me permission to use the psychology students as subjects.

were obtained from the head of the course. Grades in physics had been recorded and given to the students in the form of numbers from zero to 100. Accordingly, it was possible to use these numbers without recourse to translation.

## Results

*Relation between aspiration and achievement.*—Besides the levels of aspiration and of subsequent achievement (the examination grade), mean grades on preceding examinations were calculated for all groups, to give the level of past achievement. Tables I, II, and III present the zero-order product-moment correlations between these three variables for the three groups.

TABLE I

Intercorrelations between Achievement Variables, Group G

| | 2. Level of Aspiration | 3. Level of Past Achievement |
|---|---|---|
| 1. Grade received on examination | $r_{12} = .40$ | $r_{13} = .72$ |
| 2. Level of aspiration | — | $r_{23} = .54$ |

$N = 33$; an $r$ of .34 is significant at the 5 percent level, .44 at the 1 percent level.

TABLE II

Intercorrelations between Achievement Variables, Group E

| | 2. Level of Aspiration | 3. Level of Past Achievement |
|---|---|---|
| 1. Grade received on examination | $r_{12} = .42$ | $r_{13} = .81$ |
| 2. Level of aspiration | — | $r_{23} = .59$ |

$N = 37$; an $r$ of .32 is significant at the 5 percent level, .42 at the 1 percent level.

TABLE III

Intercorrelations between Achievement Variables, Group C

| | 2. Level of Aspiration | 3. Level of Past Achievement |
|---|---|---|
| 1. Grade received on examination | $r_{12} = .50$ | $r_{13} = .79$ |
| 2. Level of aspiration | — | $r_{23} = .57$ |

$N = 98$; for $N = 100$, an $r$ of .197 is significant at the 5 percent level, .256 at the 1 percent level.

In all groups, there is a marked positive relation between level of aspiration and subsequent achievement. Taken in isolation, this finding would tend to bear out the notion that a high aspiration represented a real goal that caused the *S* to work harder and achieve more. The higher correlations between the other two pairs of variables suggest that this conclusion is unwarranted. They tell us that examination grades may be predicted pretty well from past performance, and that both goals and expectations are based very largely on the level of previous achievement.

It is not surprising, then, to find that the correlation between aspiration and later performance drops to zero when ability (as measured by past performance) is held constant through partial correlation. This finding holds true for all groups:

Group E: $r_{12.3} = -.11$ Group G: $r_{12.3} = .02$ Group C: $r_{12.3} = .10$

It will be noted that the levels of aspiration have been used directly in this experiment, without being converted into D-scores, as is usual. One may correct an aspiration for the preceding level of achievement on which it is based by subtraction (the D-score technique); or one may hold the level of achievement constant by means of partial correlation. Though partial correlation is the preferable method (see **20**), there is evidence that in the present experiment the two methods give similar results. For Group G, D-scores were computed by subtracting the mean grade on preceding examinations from the level of aspiration for each *S*. When they are correlated with subsequent achievement, the result is the same as that obtained by partial correlation:

$$r_{1D} = .05 \qquad r_{12.3} = .02$$

All of the above results tend to verify the prediction of the defense-hypothesis, that there would be no correlation between aspiration and achievement with ability held constant, and they offer no support to the prediction of the incentive-hypothesis, that there would be appreciable positive correlation.[7]

*Differences between goals and expectations.*—Tables I and II, and the similar partial correlations from the two sets of data, argue that there is little functional difference between levels of aspiration stated in terms of goals or of expectations. The hypothesis of Bayton and Irwin, that more incentive-effect would be found if goals were used rather than expectations, must be rejected on the basis of these data. It is true, there is less positive relation between aspiration and achievement with ability held constant when *S*s are asked what they expect (Group E) than when they are asked what they are trying for. All of the partial correlations are close to zero, however, and the slight negative correlation in Group E is insignificantly different from the larger and more reliable of the other two positive ones (CR of $z = 1.06$).

[7] It was thought that the variability of past achievement might be related to level of aspiration, on the supposition that a person might feel that he had a chance for a higher grade if his previous grades had varied a good deal. For Group G the ranges of past grades, and for Group C their variances were computed, correlated with the other variables, and second-order partial correlation coefficients were calculated. The values finally arrived at were slightly closer to zero than the first-order correlations, but the effect was considered negligible, and so the extra variable was not used with Group E. For Group G, $r_{12.34} = .02$; for Group C, $r_{12.34} = .09$.

Differences of another kind between the two types of aspirations were both large and reliable. Mean past achievement levels were subtracted from the levels of aspiration given by Groups E and G, and the 70 resulting D-scores were subjected to an analysis of variance. Table IV presents a summary of this analysis.

TABLE IV

ANALYSIS OF VARIANCE OF D-SCORES ACCORDING TO INSTRUCTIONS AND SCHOOLS, GROUPS G AND E COMBINED

| | Harvard | | Radcliffe | | Total | |
|---|---|---|---|---|---|---|
| | Mean D-score | *N* | Mean D-score | *N* | Mean D-score | *N* |
| Expectations | .75 | 10 | 8.52 | 27 | 6.42 | 37 |
| Goals | 11.25 | 8 | 15.60 | 25 | 14.54 | 33 |
| Total | 5.42 | 18 | 11.92 | 52 | 10.25 | 70 |

$\Sigma X = 7175$ $\Sigma X^2 = 12831.250$ $NM^2 = 7354.375$

| Source of Variance | Sum of Squares | d.f. | $s^2$ | *F* | *P* |
|---|---|---|---|---|---|
| Instructions | 1151.94 | 1 | 1151.94 | 19.25 | <.01 |
| Schools | 566.06 | 1 | 566.06 | 9.46 | <.01 |
| Interaction | 10.99 | 4 | 2.75 | | |
| Within cells | 3747.87 | 63 | 59.49 | | |
| Total | 5476.86 | 69 | | | |

For 60 d.f. × 1 d.f., when *F* is 7.08 or larger, $P = .01$.

The mean of the *goal* D-scores is more than twice as large as that of the *expectations*. So large a difference would occur by sampling error alone much less than once in 100 samples of this size. Obviously, expectations are more realistically related to the past, while goals are more uninhibitedly reflective of wish.

As a check on the validity of the numerical transformation of the letter grades in terms of which aspirations and achievements were given, the same null hypothesis of no difference between D-scores of groups E and G was tested by the chi-square method. The *S*s in each group were divided into three classes. In the negative-and-zero class was put everyone whose level of aspiration was a lower letter grade than his previous highest grade, or the same. In the low-positive class went all those whose levels of aspiration were higher than their previous best, but not as much higher as one full grade. In the high-positive class were put *S*s whose levels of aspiration were a full grade higher than their previous best, or more. Table V shows that positive D-scores, particularly high ones, are significantly more likely to occur with goals than with expectations, and confirms the analysis of variance.

In Table IV there may also be found a somewhat unexpected result, that the Radcliffe students' D-scores are twice as large as the Harvard students', and that the difference is statistically reliable. This finding holds true for both Goal and Expectation groups, and the insignificance of the Interactions 'mean square' ($s^2$) shows that the difference is uniform. Now, there is a sizeable negative correla-

TABLE V

DIFFERENCES IN D-SCORES BASED ON EXPECTATIONS AND ON GOALS

| Sign and Size of D-scores | Number of Aspirations Given in Terms of: | | Total | Chi-square |
|---|---|---|---|---|
| | Expectations | Goals | | |
| Negative and zero..... | 27 | 6 | 33 | 11.114 |
| Low positive..... | 7 | 12 | 19 | 1.965 |
| High positive..... | 3 | 15 | 18 | 9.448 |
| Total..... | 37 | 33 | 70 | $\chi^2 = 22.527$ |

When chi-square = 18.5 at two degrees of freedom, $P = .0001$.

tion between achievement level and D-score (in Group G, the $r$ is $-.41$). There might be a partial explanation of the difference between schools, therefore, if the Radcliffe achievement levels were very much lower than those of Harvard *S*s, but the difference is insignificant, though it is in this direction (1.3 points; CR = .08). It would appear, then, that the women were more swayed from rational evaluation on the basis of past performance, and let their expectations and goals be guided to a greater extent by the wish to succeed. In light of the reported tendency of Radcliffe women to suffer much more from feelings of intellectual inferiority than do Harvard men (1, p. 174), this hypothesis seems tenable. The result is at variance with incidental findings in other experiments (2, 7) that women tend to have lower D-scores, a fact which suggests that the special Radcliffe-Harvard relationship may be responsible.

*Other esteem-defensive behavior.*—The *S*s in Group C were invited to write any comments they pleased on the slips of paper on which they recorded their goals. This instruction was permissive rather than directive, and no particular content for such comments was suggested. Some light is thrown on the nature of the goal-setting process by the comments, which 28 *S*s wrote.

First, a disproportionately large part of the comments were made by *S*s who gave the very highest goal: 100. Thirty percent of all *S*s reported goals of 100, but 54 percent of the comments came from this group. Chi-square for the frequency difference is 4.245 (one degree of freedom), which represents a probability of only .02 to .05 that the

apparent greater frequency of comments by the 100-goal group is attributable to sampling error. [It is apparent that many of the *S*s felt under some pressure to say that their goal was perfection—there were four who gave other goals but wrote comments explaining why they had *not* chosen 100—and also that the unrealistic nature of such a goal caused many of them to feel a need to justify their action when they chose it.]

Second, among the group of *S*s who wanted 100, those who made comments had significantly *poorer* past performance records than those who made no comments; in other words, D-scores for their aspirations were higher. The difference in mean achievement level is 10.7 points; $t$ is 3.29, when (for 28 degrees of freedom) a $t$ of 2.76 is significant at the one percent level. This result clarifies the nature of the conflict somewhat. [Certain *S*s are impelled to give 100 as their goal; those whose past record shows them least able to approach perfection are the ones who are forced—apparently by the need to appear realistic—to justify their aspirations.]

Third, an analysis of the comments made by the *S*s whose aspirations were least like realistic predictions, the 100-goal group, reveals a somewhat more clearly defensive quality, as compared with the total group of commentators. In the 28 comments, there are four easily distinguishable types and a residual category:

> Eight explain their goals as the score necessary to get a particular final grade in the course.
>
> For an example, see the comment previously quoted, under 'Subjects.'
>
> Seven distinguish between their goals and expectations.
>
> "Naturally, anyone would *try* for 100%. I don't expect to get over 85%."
>
> "What I am trying to make and what I expect to make is two different things."
>
> Five explain why they chose the goal of 100.
>
> "One tries to do his best; therefore, he aims at 100, because owing to his self-confidence, he considers this 'perfection' his best performance."
>
> "I think that anyone who wants to get a good grade on a test will try for 100, even though he's fairly sure he can't get it."
>
> Two extrapunitively call the investigation silly.
>
> "I always try to do the best I can, it is impossible to say I want one definite grade. This questionnaire is silly."
>
> Six made miscellaneous remarks, no two of which were alike, but all of which were comments on the grade chosen or reasons for selecting it.

One *S* who put down 85 said, "I chose the above as a *practical* goal."

All of the 12 *S*s who wrote explanations of the second two types gave 100 as their goal; another of the 15 commentators with that goal made the aggressive remark quoted above; one was unclassified; and only one fell in the least defensive category, the first.

It cannot be claimed that the evidence from the incidental comments of a minority of the *S*s establishes the esteem-defensive nature of the levels of aspiration, but it is consistent with such an explanation and at least argues in its favor. It seems plain that the demand for a self-evaluation of performance in an activity that is important for self-esteem put these *S*s into a conflict situation, in which one element was the opinion the *E* would have of them. The particular level chosen is then part of an organized attempt to keep esteem high. If giving a level of aspiration were merely a matter of symbolic representation of the strength of motivation, it is hard to see why there would be any need for face-saving on the part of so many *S*s.

## Discussion

On the basis of the results that have been set forth here, the incentive hypothesis has not yet been cleanly refuted. It is true that the statistical removal of the effect of past performance level reduces the correlation between aspiration and achievement to insignificance. There are two steps along the deductive path from this finding to either of the original hypotheses, however. One of them, not yet examined, is the assumption that a higher intensity of motivation will manifest itself in increased diligence, which will in turn be rewarded by a better grade. It will not profit us much to challenge the assumption that earnest application of a high degree of intelligence results in superior performance on an examination. But is it safe to say that the appropriate kind of effort is most likely to appear when motivation is high? It is true that this is an accepted principle of pedagogy; but it is just as true that case after case could be adduced in which *excessive* ambition and the accompanying anxiety over achievement leads to disruption of performance—the phenomenon of 'trying too hard.' Even so, however, one would expect more than a random relation between motivation and achievement (always holding ability constant), and there is no evidence of a curvilinear relationship in the present data. There seems to be no good reason to suppose that subsequent achievement was not an adequate operational definition of motivation in the three groups of subjects, when the effect of ability had been removed.

It is possible also to argue that level of previous performance is not

a valid measure of ability, so that the desired constancy has not been achieved. Granted that the only ability an examination measures may be the ability to get good grades on examinations of a similar sort, is that not exactly the ability that we wish to control in an experiment of this kind?

It could be argued further that levels of aspiration may accurately measure motivation at one moment, but in the investigations reported here so much time elapsed between the statement and the performance that motivation might have changed in enough cases to destroy the correlation. It is true, the lapse of two or three days allows such attenuating effects to come in, though it might be pointed out that much of the effective effort devoted to achievement in an examination must come during the period of days preceding the test itself. Nevertheless, arguments such as this one can only explain why less than perfect results might be expected; when correlations very close to zero are in question, they do not help much with the burden of proof which is still upon those who claim a positive relation.

Grant that, other things being equal, a person who is strongly motivated to do well in a task should be trying for a higher level of attainment than a person not so highly motivated; might he not therefore be expected to give a higher level of aspiration if asked? It is hard to answer yes, for there is still a perilous passage between the hypothetical inner goal and the announced level of aspiration to which a person must more or less publicly commit himself. It is here that it is possible to argue on theoretical grounds that levels of aspiration must have some defensive significance. In this achievement-minded culture of ours, announced goals or expectations about performance are to some extent self-evaluations, and they are matters about which the mores have a good deal to say. [A person asked for a level of aspiration is under (at least partly introjected) social demands of a conflicting nature: his statements should be *realistic*, expressive of enough *modesty*, and yet should show him to be properly *ambitious*.

Therefore, a person is doubly in peril of losing esteem from himself and from his fellows when the demand for a level of aspiration is put to him.] He knows that his subsequent performance will be judged and that, no matter what its nature, at least some slight relevance to his personal worth will be found. An important part of the frame of reference in which the judgment will be made is the stated level of aspiration. Moreover, the level of aspiration itself is something that occasions social evaluation, quite apart from what follows it. It seems reasonable to assume that all of the verbal behavior of the person thus 'put on the spot' is involved to some extent in a patterned attempt to deal with the threat to esteem. Evidence from the comments of Group C supports this assumption, but in an experiment on

large groups in classrooms or lecture halls, it is obviously impossible to observe with sufficient accuracy enough behavior to get more than the scantiest idea of what such patterns might be.[8]

If such an interpretation is accepted, the role of ego-involvement is highlighted. Bayton found some positive relation between levels of aspiration and achievement in the more ego-involved of two tasks, with previous performance held constant. Yet when, in line with his suggestion, ego-involvement was increased considerably in this experiment, the relation fell back to zero. I have been able to reconcile his results and mine only by the hypothesis of a curvilinear relation between the degree of ego-involvement in a task and the extent to which levels of aspiration measure achievant motivation. The hypothesis may be stated in the form of three propositions.

(1) *When ego-involvement is minimal, levels of aspiration have little motivational significance, being primarily rational judgments.* When Bayton's *S*s were working on the cancellation test, it seems reasonable to suppose that their ego-involvement was very slight indeed, because of their own statements, and because of the meaninglessness of the task, as presented. Therefore, no significant relation between aspiration and later achievement appeared. Differences in D-scores from different *S*s did occur, but may be attributed to momentary situational influences of a trivial nature, and to differences in forces induced by the *E*; on the whole, however, the estimates should have been primarily realistic judgments.

In support of this view is the finding of Frank that expectations followed announced achievement very closely, giving D-scores consistently close to zero, when *S*s were read a series of scores ostensibly attained by an unidentified person and were asked to predict what the next score would be. In contrast, predictions were higher and more rigid—less like realistic extrapolations—when the same scores were given to the *S*s as their own after each trial in a test (7). McGehee reports very similar findings in a comparison of estimates made by *S*s for themselves in dart-throwing, and estimates made for someone else (19).

(2) *When ego-involvement is present but at low intensities, levels of aspiration have little defensive meaning, but reflect to some extent the intensity of motivation.* If there is a chance for moderate gains or losses of self-esteem, realistic prediction becomes somewhat subordinated to wishful thinking on the part of those persons who have the strongest wishes, but the danger to esteem is not great enough to arouse defensive measures at the time of the aspiration-giving. The

[8] In a small-scale experiment with an intensively studied group of college *S*s, I have found a dozen and a half patterns of esteem-defense into which measures of level of aspiration and other self-evaluations fit, patterns that are highly correlated with personality characteristics (12).

direct expression of the strength of motivation in the height of announced goals or expectations will not be much obscured by the devious strategies of esteem-defense, when all *S*s are operating within a certain range of ego-involvement. Bayton's results with his more-involved group would be explained by this part of the hypothesis.

[(3) *When ego-involvement surpasses a certain limit, defensive considerations become paramount, and the level of aspiration becomes more complexly determined.*] In part a rational prediction, influenced by the strength of motivation, finally it is subjected to the more important dictates of the person's habitual modes of esteem-defense. The correlation between aspiration and achievement, independent of ability, falls to zero, as in the experimental results reported here.

A pair of findings from an experiment which I have reported elsewhere (**13**) supports these contentions, though the results are not quite comparable. The *S*s were given tests of two abilities after it had been determined in each case which was the more ego-involved. In tests of the less-involved ability, there were positive correlations of .64 and .25 between aspirations corrected for past achievement, and *improvement* (not absolute level) of performance. *S*s were moderately involved in these tests, but very highly so in the other group of tests, where the correlation between corrected aspirations and improvement of performance fell to the negative values of $-.76$ and $-.62$ ($N = 13$). Measures of both aspiration and achievement were somewhat different from the ones in the present experiment, but the results are suggestive in showing a reversal of the relation with an increase of ego-involvement.

The fact that diverse experimental results can be fitted into a rather complicated hypothesis does not prove it, of course. Neither can it be said that the defense-hypothesis with which this investigation started has been brought as near to proof as the incentive-hypothesis has to disproof. Nevertheless, the defense-hypothesis, as modified to take into consideration the factor of ego-involvement, does unify into a consistent framework all of the reported experimental work on the relation between aspirations and performance, and facts from somewhat different investigations as well.

## Summary

Levels of aspiration were obtained from 168 undergraduates who were about to take examinations in college courses. Some *S*s were asked for expectations, others for goals. Deductions about the relation between level of aspiration and subsequent achievement were made on the basis of two prevailing conceptions of the level of aspiration. From the incentive-hypothesis, that levels of aspiration

directly reflect the degree of a person's motivation in the task being performed, it was predicted that there would be a positive correlation between aspirations and excellence of following performance, when the level of previous achievement was held constant. From the defense-hypothesis, that levels of aspiration are to be understood primarily as part of the defensive behavior of a person whose self-esteem is threatened, it was predicted that no such correlation would be found.

Partial correlations between aspirations and examination grades were computed, with past level of attainment held constant. Results were as follows:

1. In the present group of Ss, no correlation between aspiration and achievement existed, when past level of achievement was partialled out.

2. This lack of relationship was equally true whether aspirations were obtained in terms of goals or expectations, and whether the examinations were in psychology or in physics.

3. Evidence from written comments made by part of one group of Ss indicates that the more aspirations diverge from realistic predictions, the more defensive the behavior of the S becomes.

4. D-scores based upon goals were significantly larger than D-scores based upon expectations; D-scores of the male and of the female students also differed significantly.

Results were reconciled with other experimental findings in terms of a hypothetical curvilinear relation between degree of ego-involvement and the degree of correlation between aspiration and achievement.

## References


1. Allport, G. W. *Personality, a psychological interpretation.* New York: Holt, 1937.
2. Anderson, H. H., & Brandt, H. F. Study of motivation involving self-announced goals of fifth grade children and the concept of level of aspiration. *J. soc. Psychol.*, 1939, **10**, 209–232.
3. Bayton, J. A. Interrelations between levels of aspiration, performance, and estimates of past performance. *J. exp. Psychol.*, 1943, **33**, 1–21.
4. Festinger, L. Wish, expectation, and group standards as factors influencing levels of aspiration. *J. abnorm. soc. Psychol.*, 1942, **37**, 184–200.
5. Filter, R. O. Estimates of the amount of work one can do. *J. appl. Psychol.*, 1927, **11**, 58–67.
6. Frank, J. D. Individual differences in certain aspects of the level of aspiration. *Amer. J. Psychol.*, 1935, **47**, 119–128.
7. Frank, J. D. A comparison between certain properties of the level of aspiration and random guessing. *J. Psychol.*, 1936, **3**, 43–62.
8. Freud, A. *The ego and the mechanisms of defense.* London: Hogarth Press, 1937.
9. Gardner, J. W. Level of aspiration in response to a prearranged sequence of scores. *J. exp. Psychol.*, 1939, **25**, 601–621.
10. Gardner, J. W. The use of the term 'level of aspiration.' *Psychol. Rev.*, 1940, **47**, 59–68.

11. GOULD, R. An experimental analysis of 'level of aspiration.' *Genet. Psychol. Monogr.*, 1939, **21**, 3–115.
12. HOLT, R. R. Motivational factors in levels of aspiration. Ph.D. dissertation, Harvard University, 1943.
13. HOLT, R. R. Effects of ego-involvement upon levels of aspiration. *Psychiatry*, 1945, 8, 299–317.
14. HOPPE, F. Erfolg und Misserfolg. *Psychol. Forsch.*, 1930, **14**, 1–62.
15. IRWIN, F. W., & MINTZER, M. G. Effect of differences in instructions and motivation upon measures of the level of aspiration. *Amer. J. Psychol.*, 1942, **55**, 400–406.
16. JUCKNAT, M. Leistung, Anspruchsniveau und Selbstbewusstsein. *Psychol. Forsch.*, 1937, **22**, 89–179.
17. KNEELAND, N. Self-estimates of improvement in repeated tasks. *Arch. Psychol.*, 1934, **25**, No. 163, 1–75.
18. MAY, M. A., & DOOB, L. W. Competition and cooperation. *Soc. Sci. Res. Coun. Bull.*, 1937, No. 25, 1–191.
19. MCGEHEE, W. Judgment and the level of aspiration. *J. gen. Psychol.*, 1940, **22**, 3–15.
20. PRESTON, M. G. Use of the coefficient of correlation in the study of the D-score for the level of aspiration. *Amer. J. Psychol.*, 1942, **55**, 442–446.
21. ROTTER, J. S. Level of aspiration as a method of studying personality. I. A critical review of methodology. *Psychol. Rev.*, 1942, **49**, 463–474.
22. SEARS, P. S. Level of aspiration in relation to some variables of personality: clinical studies. *J. soc. Psychol.*, 1941, **14**, 311–336.

# THE EFFECT OF CONTEXT UPON JUDGMENTS OF PITCH DIFFERENCES

BY T. KOESTER AND W. N. SCHOENFELD

*Columbia University*

When stimuli distributed along some dimension are presented interchangeably for judgment in the same sitting, it has been found that constant errors show a systematic trend from positive to negative (or toward increasing negativity when all the errors are negative in sign) with increasing value of the stimuli. Graphical representations of this type of stimulus-context phenomenon will be referred to in the present study as 'negatively directed constant error functions' in order to avoid terminology ('central tendency of judgment,' 'assimilation,' etc.) weighted with previous theoretical argument.

Representative of the studies in which negatively directed CE-functions have been found are those of Hollingworth (**3**, **4**) on judgments of linear extent and sizes of squares, Woodrow (**11**) on weight-lifting, and Needham (**8**) on loudness differences. Hollingworth regarded the negatively directed CE-function as illustrative of a 'central tendency of judgment' in which standards greater than a norm (determined either by *O*'s previous experience or in the course of an experimental sitting) are underestimated and standards smaller than the norm are overestimated. He writes (**3**, p. 463): "in all estimates of stimuli belonging to a given range or group we tend to form our judgments around the median value of the series—toward this mean each judgment is shifted by virtue of a mental set corresponding to the particular range in question."

Woodrow was of the opinion that when a variety of standards, differing by relatively small amounts, are used interchangeably during a sitting, observers become 'set' for or adjusted to the weighted average of all the interacting stimulus values, greater weights being assigned to the most recent stimuli in a series of presentations and the greatest weight to the standard immediately preceding the judgment about to be made. This set acts in such a way as to assimilate to itself the impressions of the standards differing from it, thus diminishing the apparent differences among the standards and increasing the probability that the comparison stimuli of standards greater than the mean of the series will be judged 'higher' and those of standards smaller than the mean will be judged 'lower,' and consequently giving rise to a negatively directed CE-function.

It may be noted that both Hollingworth and Woodrow consider only the standard and not its comparison stimulus as being affected by context, and they suppose that changes in the standards may be measured through the unaltered impressions of the comparison stimuli with which they are paired.[1] This limitation of the context effect to the standard stimulus may be correct, but no experimental evidence can be adduced to support it. The fact is that in the comparison of two stimulus impressions, the individual stabilities of which are not verifiable by external means, it is impossible to decide whether an altered comparative judgment is due to a change in one impression or the other.

Needham's main interest in working with interacting loudness levels was in the effect which a lengthening in the time interval separating the standard and comparison stimuli would exercise upon the CE-function. His data, which showed an increasing negative slope in the CE-function (as plotted by the writers) with lengthening time interval, led him to the generalization that "as the time interval separating the members of a pair of comparison stimuli is lengthened, the 'predisposing effects' of the preceding stimulation become more and more effective" (**8**, p. 539). One of Hollingworth's observers showed a similar trend with an extension of the time interval from two to 30 sec. (**4**). Likewise, in the group data presented by Lauenstein (**6**) from one of his experiments on the judgment of loudness differences, this same trend appeared for five closely spaced levels as time intervals increased from .20 to 5 sec., but was reversed between 10 and 20 sec. The latter's finding may require qualification, however, since the frequency of the equal judgments fell markedly with increasing time intervals and the possibility remains that two-category data may not have revealed a different trend for the shorter intervals.

The principal aim of the present study was to determine whether the negatively directed CE-function typically reported for intensitive and extensitive judgments would also occur in the case of qualitative judgments made in a context of two or more interacting pitch stimuli. Data presented by Wolfe (**10**) in his study of pitch discrimination indicated that judgments of qualities would prove to be no exception. In one of his experiments in which 11 standards, ranging from 144 to 1004 cycles, were employed interchangeably, the combined data for all of his observers and for a number of time intervals between standard and comparison stimuli ranging from 1 to 60 sec. showed a tendency for positive constant errors to appear with the lower standards and negative errors with the higher. Unfortunately, his data are

[1] In his study on judgments of linear extent, Hollingworth used the method of average error. Here, the reproduction of the standard is presumably considered to be unaffected by context.

not presented in such a way as to permit an examination of the effect of the separate time intervals upon the final result. Wolfe interpreted his findings as indicating that stimulus level alone, apart from any interaction effects, was a significant variable in determining the sign of the constant error. This conclusion may be questioned, however, in the light of the evidence recently reported by Koester (5) that there is no systematic trend in the direction or magnitude of the constant error over a frequency range from 200 to 4000 cycles when only a single stimulus level is used in any one sitting.

Secondary aims of the present study concerned the following points:

1. The importance of the differences separating the standards employed in one sitting. Woodrow (11) has expressed doubt that the negatively directed CE-function would have occurred in his study if the standards had varied from 20 to 600 gm. instead of from 110 to 200 gm. "It seems inevitable," he writes, "that when the range of stimuli is very great, a stimulus near either extreme of this range would result in an exception to the rule that *S* tends to adapt to the average of the entire series" (11, p. 413). No evidence was produced, however, to support this contention.

2. The number of standards necessary to obtain interaction effects. The minimum number of standards employed in previous studies has been four.

3. The importance of the immediately preceding stimuli.

4. The effect of extensions in the time interval separating standard and comparison pitch stimuli.

5. A comparison of the performance of naive and practised observers. From preliminary work, it seemed that some difference might be expected.

## Experimental Findings

### *Part I*

The three experiments to be described here were intended to yield information on the problems listed above in the case of naive observers (data for practised *O*s will be presented in Part II). These observers were 15 female undergraduate students with no experience in making psychophysical judgments and uninformed as to the purposes of these experiments. All stimuli were presented by means of a twin-oscillator designed and described by Stevens and Gerbrands (9) under technical arrangements outlined in an earlier study (5).

*Procedure A.*—Pairs of moderately loud stimuli were presented for comparison judgments to seven *O*s at five moderately loud frequency levels: 1000, 1030, 1060, 1090, and 1120 cycles. Two comparison stimuli were 12 cycles above and below each of the standards,[2] and the third

[2] In a study by Koester (5) most of a group of 14 naive *O*s made 100 percent correct judgments on a 12-cycle frequency difference. This large difference was employed here since preliminary data confirmed that obtained by previous investigators showing that discrimination is more difficult when several standards are used interchangeably at the same sitting.

was the standard repeated. These comparison stimuli were presented for judgment with equal frequency. The standards were given in a random order with no repetition of any one standard until all five had been given. The time interval between pairs of standard and comparison stimuli was eight sec., and the intra-pair interval was one sec. for one-half the judgments made by each *O* and six sec. for the other half. Each *O* made a total of 240 judgments (24 per time interval at each frequency level) in four blocks of 60 judgments each, with a rest pause inserted after each block. The intra-pair time interval order for the four blocks was 1–6–6–1 for four *O*s and 6–1–1–6 for three *O*s. Judgments were reported in two categories, 'higher' and 'lower.'

*Procedure B.*—Four other *O*s were tested with only two standards presented alternately. Three pairs of standards were employed: (*a*) 1000–1010, (*b*) 1000–1100, (*c*) 1000–2000 cycles. All judgments were taken on one pair of standards before proceeding to the next, and each *O* completed his work on the three pairs of standards in one sitting. The order of presenting the pairs was a–b–c for two *O*s and c–b–a for the other two *O*s. Each *O* made 48 judgments on each pair of standards in an uninterrupted series using the two categories 'higher' and 'lower,' and, following a rest, went on to the next pair of standards. The comparison stimuli for the 1000,

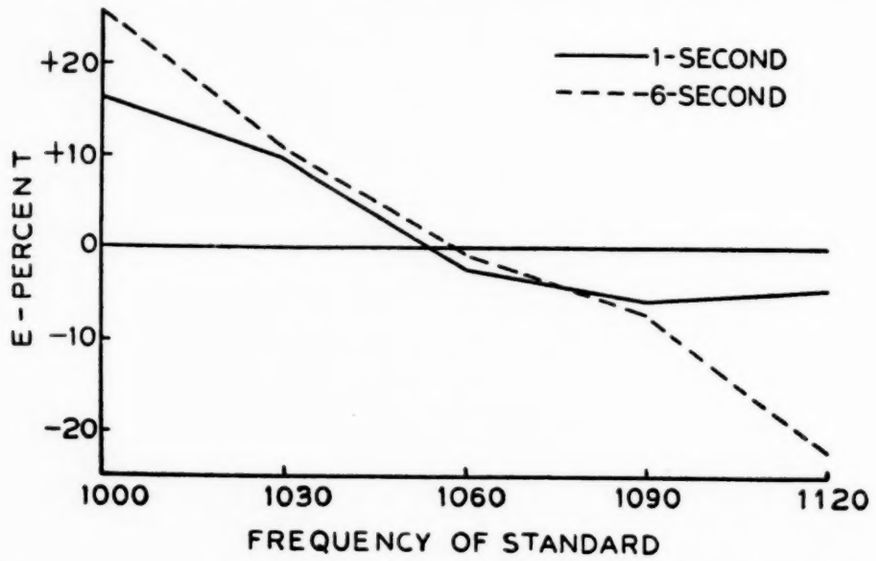


FIG. 1. The constant error (E-percent) for pitch judgments made by seven *O*s at five frequency levels and at one- and six-sec. time intervals. Data in Table I.

1010, and 1100-cycle standards were ±12 cycles, and for 2000 cycles they were ±24 cycles, while the standard repeated gave a third comparison stimulus; all comparison stimuli were used with equal frequency. The time interval between standard and comparison stimuli was six sec., while eight sec. separated one standard-comparison stimulus pair from the next.

*Procedure C.*—Four new *O*s were tested with only two standards, 1000 and 1100 cycles, but three successive judgments were obtained with one standard before passing to the other. The three comparison stimuli were ±12 cycles and the repeated standard. The order of presenting comparison stimuli was determined as follows: from 9 cards, 3 of which were marked 'higher,' 3 'equal,' and 3 'lower,' a random order was made separately for all of the first, then the second, and finally the third presentations following stimulation at the preceding level. Each *O* made a total of 72 judgments, either 'higher' or 'lower,' for each standard in one sitting, with rest pauses allowed after every 36 judgments. Six sec. separated a standard and comparison stimulus, and the interval between standard-comparison pairs was eight sec.

## *Results*

The pitch judgments of naive *O*s made in a 5-standard context yield a negatively directed CE-function [3] (Fig. 1 and Table I; data

[3] E-percent (see footnote, Table I) is used as the measure of the constant error in the present study.

TABLE I

DISTRIBUTION AND E-PERCENT VALUES FOR PITCH JUDGMENTS MADE BY SEVEN UNPRACTISED *O*s AT FIVE FREQUENCY LEVELS AND TWO TIME INTERVALS

| Frequency Level in Cycles | One-Sec. Interval | | | Six-Sec. Interval | | |
|---|---|---|---|---|---|---|
| | Distribution | | E-% | Distribution | | E-% |
| | h | l | | h | l | |
| 1000 | 57 | 111 | +16.07 | 41 | 127 | +25.59 |
| 1030 | 68 | 100 | + 9.52 | 66 | 102 | +10.71 |
| 1060 | 88 | 80 | − 2.38 | 85 | 83 | − .59 |
| 1090 | 94 | 84 | − 5.95 | 96 | 72 | − 7.14 |
| 1120 | 91 | 77 | − 4.17 | 121 | 47 | −22.02 |

E-percent = $100 \frac{L}{N} - 50$. This is a measure of the constant error (2, 5) showing the degree to which the distribution of judgments made on comparison stimuli spaced equidistantly about a standard departs from symmetry or is in excess of a chance distribution. 'L' refers to the number of 'lower' judgments and 'N' to the total number of 'higher' and 'lower' judgments. A positive E-percent indicates a preponderance of 'lower' and a negative E-percent a preponderance of 'higher' judgments. The quantities in the 'h' and 'l' columns indicate the total number of 'higher' and 'lower' judgments made by the seven *O*s. The trend for each of the individual *O*s was similar.

obtained under Procedure A) similar to that reported by previous investigators for intensive and extensive judgments. The increase in the intra-pair interval from one to six sec. accentuated the effect mainly for the standards at either extreme of the series. In Needham's (8) study on loudness, each of his four standards was influenced by the same time interval extension.

TABLE II

DISTRIBUTION AND E-PERCENT VALUES FOR PITCH JUDGMENTS MADE BY FOUR UNPRACTISED *O*s AT THREE PAIRS OF FREQUENCY LEVELS

| Pair of Standards | | $S_1$ | | | $S_2$ | | |
|---|---|---|---|---|---|---|---|
| $S_1$ | $S_2$ | Distribution | | E-% | Distribution | | E-% |
| | | h | l | | h | l | |
| 1000~ | 1010~ | 47 | 49 | + 1.04 | 56 | 40 | − 8.33 |
| 1000~ | 1100~ | 25 | 71 | +23.96 | 74 | 22 | −27.08 |
| 1000~ | 2000~ | 32 | 64 | +16.67 | 57 | 39 | − 9.37 |

The trend for the individual *O*s was similar to that shown in the combined results above.

From Table II (data from Procedure B) it may be seen that the use of only two alternating standards produces the same sort of interaction effect as when five standards are used interchangeably, that is, the constant error is positive for the lower standard and nega-

tive for the higher. No consistent trend is apparent, however, for the increasing frequency difference between the two standards, inasmuch as the effect for the 10-cycle difference is the smallest, while that for the 100-cycle difference is the greatest. Since the effect is quite strong with the latter difference which is a large one in pitch discrimination, Woodrow's (11) opinion that large differences among weight standards would eradicate the effect plainly cannot be extended to other kinds of judged material if, indeed, it is valid for weights in the first instance.

The importance to the interaction effect of the immediately preceding stimulation is evidenced by the decreasing E-percents (Table III; data from Procedure C) for both the 1000 and 1100-cycle stand-

TABLE III

DISTRIBUTION AND E-PERCENT VALUES FOR PITCH JUDGMENTS MADE BY FOUR UNPRACTISED *O*s ON THREE SUCCESSIVE PRESENTATIONS OF A STANDARD-COMPARISON PAIR OF STIMULI FOLLOWING STIMULATION AT A HIGHER OR LOWER FREQUENCY LEVEL

| Judgment following Preceding Level | 1000 ~ Level | | | 1100 ~ Level | | |
|---|---|---|---|---|---|---|
| | Distribution | | E-% | Distribution | | |
| | h | l | | h | l | |
| First..... | 20 | 76 | +29.17 | 83 | 13 | − |
| Second..... | 37 | 59 | +11.46 | 73 | 23 | −[illegible] |
| Third..... | 35 | 61 | +13.54 | 66 | 30 | − |

Individual and combined trends were similar. The Table reads as follows: e.g., th[illegible] 20 'h' and 76 'l' judgments at the 1000-cycle level when the immediately preceding judgm[illegible] been made at the 1100-cycle level; there were 66 'h' and 30 'l' judgments at the 1100-cy[illegible] level when the two immediately preceding had also been made at the 1100-cycle level, etc.

ards with three successive presentations of each. The considerable effect still remaining, however, on the third judgment made at the same frequency level following stimulation at either the higher or lower frequency level is an indication that other factors are operative.

## *Part II*

In the following experiments practised observers were tested under experimental conditions similar to those used in Procedures A and B of Part I with naive observers. In addition, a more extensive range of time intervals between standard and comparison stimuli was employed.

*Procedure A.*—Four highly practised *O*s made judgments under the same conditions as outlined in Procedures A and B of Part I with the exceptions (1) that the comparison stimuli were adjusted to the known discriminative capacities of the *O*s, while still being about twice as great as the *O*s detected correctly 75 percent of the time when the standards were presented

separately,[4] and (2) that three categories of judgment ('higher,' 'equal,' and 'lower') were permitted instead of two. Two *O*s, TK and WS (the writers), were familiar with the purpose of the experiment, while two, EK and BW, were not.

*Procedure B.*—The observers were three (WS, TK, EK) of the four who served in Procedure A. Standard and comparison stimuli at two frequency levels, 1000 and 1100 cycles, employed interchangeably, were presented with four different intra-pair time intervals: 0, 1, 6, 12 sec. There were three comparison stimuli: one higher, one correspondingly lower, and the third equal to the standard. The higher and lower comparison stimuli were selected to give differences about twice as large as those which could be detected correctly 75 percent of the time when the standards were presented separately. Four sittings were required of each observer, and 120 judgments were taken at a single sitting, 30 at each time interval. All judgments were made at one time interval before proceeding to the next, the time interval order being 0–1–6–12 at sittings 1 and 3 and 12–6–1–0 at sittings 2 and 4. Again three categories of judgment were allowed.

Two new *O*s who had not served before in any of the experiments made judgments at only the 6-sec. time interval on four successive days. They had had some practise in making psychophysical judgments, including those of pitch differences, but were not familiar with the purpose of this experiment. At each sitting 120 judgments were made. The fourth sitting of these two *O*s differed from the earlier three in that the first 90 judgments were each immediately followed by information as to correctness and by immediate repetition of a standard-comparison stimulus pair when the judgment was incorrect; following this a test series of 30 judgments without information was taken.

## *Result*

The results for *O*s practised in pitch discrimination do not show much agreement with those reported in Part I for the naive *O*s. ents made by BW on five standards at the 6-sec. interval IV: Procedure A) were the only ones which showed the nega-

TABLE IV

UTION AND E-PERCENT VALUES FOR PITCH JUDGMENTS MADE BY FOUR PRACTISED *O*s AT FIVE FREQUENCY LEVELS AND TWO TIME INTERVALS

| Time Interval | *O*s | | Frequency Level 1000~ h | e | l | 1030~ h | e | l | 1060~ h | e | l | 1090~ h | e | l | 1120~ h | e | l |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| One sec. | WS | | 11 | 9 | 4 | 7 | 3 | 14 | 10 | 9 | 5 | 12 | 7 | 5 | 9 | 0 | 15 |
| | | E-% | | −14.58 | | | +14.58 | | | −10.42 | | | −14.58 | | | +12.50 | |
| | TK | | 18 | 3 | 3 | 8 | 4 | 12 | 8 | 9 | 7 | 9 | 4 | 11 | 5 | 6 | 13 |
| | | E-% | | −31.25 | | | + 8.33 | | | − 2.08 | | | + 4.17 | | | +16.67 | |
| | EK | | 17 | 2 | 5 | 14 | 2 | 8 | 14 | 2 | 8 | 10 | 3 | 11 | 3 | 1 | 20 |
| | | E-% | | −25.00 | | | −12.50 | | | −12.50 | | | + 2.08 | | | +35.42 | |
| | BW | | 9 | 8 | 7 | 7 | 7 | 10 | 9 | 3 | 12 | 10 | 2 | 12 | 9 | 8 | 7 |
| | | E-% | | − 4.17 | | | + 6.25 | | | + 6.25 | | | + 4.17 | | | − 4.17 | |
| Six sec. | WS | | 5 | 10 | 9 | 5 | 8 | 11 | 13 | 4 | 7 | 8 | 6 | 10 | 15 | 7 | 2 |
| | | E-% | | + 8.33 | | | +12.50 | | | −12.50 | | | + 4.17 | | | −27.08 | |
| | TK | | 12 | 7 | 5 | 9 | 5 | 10 | 3 | 12 | 9 | 10 | 5 | 9 | 14 | 6 | 4 |
| | | E-% | | −14.58 | | | + 2.08 | | | +12.50 | | | − 2.08 | | | −20.83 | |
| | EK | | 16 | 6 | 2 | 19 | 1 | 4 | 15 | 2 | 7 | 16 | 2 | 6 | 12 | 4 | 8 |
| | | E-% | | −29.17 | | | −31.25 | | | −16.67 | | | −20.83 | | | − 8.33 | |
| | BW | | 3 | 11 | 10 | 10 | 2 | 12 | 7 | 0 | 17 | 14 | 0 | 10 | 15 | 4 | 5 |
| | | E-% | | +14.58 | | | + 4.17 | | | +20.83 | | | − 8.33 | | | −20.83 | |

[4] Reasons for the use of large frequency differences are given in footnote 2.

tively directed constant error trend. The results for WS, TK, EK were irregular at the 6-sec. interval; but at one sec., the constant errors for the judgments made by these *O*s on the two extreme standards of the 5-standard context were directly opposite in sign to those for the naive *O*s, being negative for the lowest and positive for the highest standard. The latter trend was also shown by these three *O*s when they were tested with two standards differing by 100 and 1000 cycles (Table V; Procedure A), and by EK when the frequency difference was 10 cycles.

Aside from the direction in the constant error trend, Table V shows that the influence of context is still present with a 1000-cycle

TABLE V

DISTRIBUTION AND E-PERCENT VALUES FOR PITCH JUDGMENTS MADE BY FOUR PRACTISED *O*s AT THREE PAIRS OF FREQUENCY LEVELS

| Pair of Standards | | *O*s | $S_1$ | | | | $S_2$ | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| | | | Distribution | | | E-% | Distribution | | | E-% |
| $S_1$ | $S_2$ | | h | e | l | | h | e | l | |
| 1000~–1010~ | | WS | 7 | 0 | 17 | +20.83 | 13 | 0 | 11 | − 4.17 |
| | | TK | 9 | 9 | 6 | − 6.25 | 14 | 7 | 3 | −22.92 |
| | | EK | 20 | 1 | 3 | −35.42 | 13 | 0 | 11 | − 4.17 |
| | | BW | 7 | 1 | 16 | +18.75 | 14 | 2 | 8 | −12.50 |
| 1000~–1100~ | | WS | 14 | 0 | 10 | − 8.33 | 12 | 0 | 12 | .00 |
| | | TK | 14 | 4 | 6 | −16.67 | 13 | 3 | 8 | −10.42 |
| | | EK | 15 | 5 | 4 | −22.92 | 11 | 0 | 13 | + 4.17 |
| | | BW | 4 | 4 | 16 | +25.00 | 16 | 1 | 7 | −18.75 |
| 1000~–2000~ | | WS | 17 | 0 | 7 | −20.83 | 9 | 0 | 15 | +12.50 |
| | | TK | 14 | 4 | 6 | −16.67 | 8 | 2 | 14 | +12.50 |
| | | EK | 21 | 0 | 3 | −37.50 | 3 | 3 | 18 | +31.25 |
| | | BW | 11 | 3 | 10 | − 2.08 | 14 | 4 | 6 | −16.67 |

separation of the standards, and to this extent there is agreement with the results for the naive *O*s (Table II).

The extension of the intra-pair time interval beyond one sec. when two standards are used alternately with practised *O*s eventuates in an increasingly negative constant error for the lower of the two standards and an increasingly positive constant error for the higher (Fig. 2 and Table VI: Procedure B). The results for the 6-sec. interval for *O*s, WS, TK, EK, are in the same direction as those reported in Table V for a similar interval and the same standards. It is only at the one-sec. interval that WS and TK show a positive constant error for the lower standard and a negative for the higher. The trend for EK is the same at all intervals. Of the two additional *O*s who made judgments at only the 6-sec. interval, MT showed a

positive constant error for the higher standard and a negative for the lower, while FA showed the reverse. At the fourth sitting, after these last two *O*s had been given information on 90 judgments, there was no change in the trend of their individual results on the critical series of 30 judgments.

## Discussion

Context in the form of either two or five pitch standards presented interchangeably for judgment in the same sitting has produced two

TABLE VI

Distribution and E-Percent Values for Pitch Judgments Made at Two Alternating Frequency Levels by Three Practised *O*s at Four Time Intervals and by Two *O*s at One Time Interval

| Time Interval in Seconds | *O*s | 1000 ~ Level | | | | 1100 ~ Level | | | |
|---|---|---|---|---|---|---|---|---|---|
| | | Distribution | | | E-% | Distribution | | | E-% |
| | | h | e | l | | h | e | l | |
| 0 | WS | 23 | 11 | 26 | + 2.50 | 15 | 27 | 18 | + 2.50 |
| | TK | 28 | 14 | 18 | − 8.33 | 21 | 21 | 18 | − 2.50 |
| | EK | 39 | 0 | 21 | −15.00 | 26 | 3 | 31 | + 4.17 |
| 1 | WS | 16 | 23 | 21 | + 4.17 | 24 | 22 | 14 | − 8.33 |
| | TK | 24 | 12 | 24 | .00 | 27 | 18 | 15 | −10.00 |
| | EK | 43 | 1 | 16 | −22.50 | 26 | 3 | 31 | + 4.17 |
| 6 | WS | 32 | 15 | 13 | −15.83 | 14 | 29 | 17 | + 2.50 |
| | TK | 28 | 19 | 13 | −12.50 | 20 | 21 | 19 | − .83 |
| | EK | 50 | 1 | 9 | −34.17 | 23 | 0 | 37 | +11.67 |
| | MT* | 109 | 39 | 32 | −21.39 | 69 | 56 | 55 | − 3.89 |
| | FA* | 55 | 42 | 83 | + 7.78 | 133 | 22 | 25 | −30.00 |
| 12 | WS | 27 | 22 | 11 | −13.33 | 11 | 28 | 21 | + 8.33 |
| | TK | 35 | 15 | 10 | −20.83 | 16 | 17 | 27 | + 9.17 |
| | EK | 51 | 2 | 7 | −36.66 | 17 | 4 | 39 | +18.33 |

* MT and FA made judgments at only the 6-sec. interval. Data reported are for their first three sittings.

opposite trends in the constant errors, thus posing a problem in explanation which previous investigators of the context effect on intensitive and extensitive stimuli, confronted with only the negatively directed constant error function, did not have to face. The findings for the naive *O*s are in agreement with those reported by earlier workers in dimensions other than pitch. The constant errors for these naive *O*s were positive for the lower stimuli and negative for the higher stimuli comprising a given context. Constant errors for some of the practised *O*s, on the contrary, were reversed in trend, being generally negative for the lower and positive for the higher

pitch standards, especially with intra-pair time intervals greater than one second. The latter type of result has not hitherto been reported.

An explanation for the reversal in the CE trend for practised *O*s may be offered in terms of a contrast between the members of a comparison pair and the immediately preceding stimulus value. Thus, if a low standard were to follow stimulation at a higher level, it would by contrast sound lower than it would out of context, with the result that its comparison stimuli, which we may assume to be less affected by contrast, since they are separated from the preceding level by a

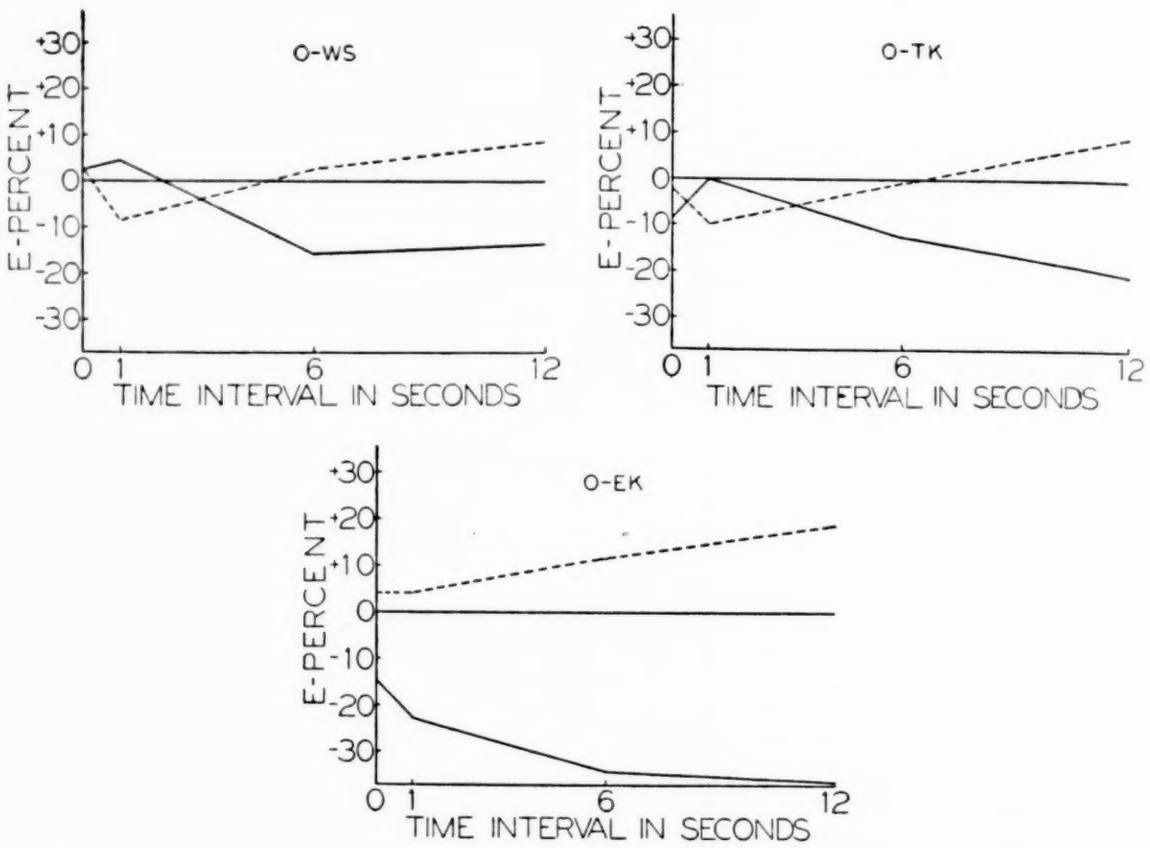


FIG. 2. The constant error (E-percent) for comparison judgments made by three practised *O*s at two frequency levels (1000 cycles, unbroken line; and 1100 cycles, dashed line) used interchangeably and at four intra-pair time intervals. Data in Table VI.

greater temporal interval and also follow their own standard rather than any other stimulation level, would more frequently sound higher. Any increase in the time interval between the standard and comparison stimuli would, by this argument, lessen the effect of contrast upon the comparison stimuli and yield increasingly negative constant errors for low standards and increasingly positive for the high. The plots in Fig. 2 bear out the latter expectation; while the importance of the preceding level of stimulation has already been pointed out (Table III). It is to be noted that, in the explanation offered here, both the standard and the comparison stimuli are regarded as being

affected by contrast, with allowance being made for differing degrees of the effect upon the two members of a comparison pair. In addition to contrast with the preceding level, some writers may prefer to consider the possibility of similar contrast also existing between the members of a standard-comparison stimulus pair and the mean of a stimulus series; but the present writers are of the opinion, that even if this were so, such contrast would make only a small contribution to the total result.

Explanations of negatively directed constant error trends in terms of assimilation and central tendency have been discussed. It was indicated that these are valid only on the assumption that the effect of context is limited to the standard or is at least greater upon the standard than upon either its reproduction or a second stimulus with which it is paired. Three additional ways of accounting, at least in part, for negatively directed constant error trends are conceivable, but logical difficulties stand in the way of accepting any one of them as a unitary explanation. First, negatively directed CE-functions would result if comparison stimuli were to be judged more often in relation to the preceding standard-comparison pair rather than to their own immediate standards. This would mean that the comparison stimuli of lower level standards, since they are preceded more often by greater than by lower stimulus values, would more frequently be judged 'lower,' and vice versa for the comparison stimuli of higher level standards. Some of the data in the present study (Table III) and in others, notably that of Fernberger (1) in weight-lifting, have shown that the preceding stimulation contributes importantly to the interaction effect. Note that, here, neither the standard nor the comparison stimuli are regarded as changing, but simply that the interaction effect is produced by the observer's referring a comparison stimulus to the preceding level of stimulation rather than to its own immediate standard. One of the difficulties with this as a unitary explanation is an equivocality similar to that involved in Hollingworth's and Woodrow's theories, since the negatively directed CE-function is as readily attributable to the assimilation of the standard to the preceding level of stimulation, with the comparison stimuli remaining unaffected, as it is to the direct referral of the comparison stimulus to the preceding level of stimulation.

A second cause of the negatively directed constant error trends with increasing value of the standard stimuli would be the disposition on the part of *O*s to make a considerable number of absolute judgments [5] either on the standard, the comparison stimulus, or on both.

[5] Martin and Mueller (7), in an extensive weight-lifting study, have shown that such judgments are made.

Familiarity with a number of stimuli in a series may result in an absolute scale in which a standard or a comparison stimulus is regarded as high or low depending upon its position in the series. Then, if a considerable number of absolute judgments are made, constant error functions would become negatively directed, because such judgments made on low stimuli would be preponderantly 'lower' and those upon the high, preponderantly 'higher.' This explanation, like the first, also has its weaknesses, but these are of a somewhat different nature. First, there is no way, aside from introspection, of knowing whether absolute judgments are made at all; second, if they are made, it is unlikely that, with *O*s instructed to make comparative judgments, they would be sufficiently numerous to produce the marked effects obtained; and third, there is no way of determining, again aside from introspection, whether the standard, the comparison stimuli, or both, are being absolutely judged.

An explanation in terms of both contrast and either assimilation, central tendency, or set, represents a third possibility. It may be argued that standards of a comparison pair contrast with varying preceding levels of stimulation, but that this effect lasts only for the duration of the standard. During the interval which separates the standard and comparison stimuli, the trace of or set for the standard is assimilated toward the mean of the series, with negatively directed CE functions resulting when the latter forces are more effective than the former and positive CE functions when contrast remains the more effective force. This explanation is more inclusive than the first two, but can hardly be offered as the only one, since here again the effect of context is limited to one member of a comparison pair.

If the above explanations and those in terms of central tendency or assimilation exhaust the possibilities, the true cause of the negatively directed constant error trends comprising a context for judgment must lie somewhere among them. No one of these may be singled out as the sole explanation, however, both because of the logical difficulties involved in so doing, and because of the likelihood that all of the factors appealed to in the various explanations, rather than any one, may be effective at some time or other within the same sitting.

There still remains the question of the directly opposite effects shown by the naive and some of the practised *O*s judging under identical stimulus-context conditions. Practise alone can hardly be singled out as the decisive variable since at least two of the practised *O*s (BW and FA) were in agreement with the naive *O*s. The discrepant findings may perhaps be attributed to the fact that through considerable practise some *O*s become less affected by certain vari-

ables or tendencies in judgment, thus permitting others, such as 'contrast,' e.g., to become the effective forces.

## Summary

The experimental findings on the judgment of pitch differences between a standard and comparison stimulus, when two or more standards are employed interchangeably at the same sitting, may be summarized as follows:

1. Constant errors for naive *O*s and some practised *O*s vary consistently from positive to negative with increasing frequency level of the standard. This is true whether two or five standards are used as a context.

2. For some practised *O*s, constant errors are negative for the lower and positive for the higher standards.

3. Lengthening intra-pair time interval is accompanied by increasing interaction or context effect for naive *O*s on the extreme standards of a 5-standard context and for practised *O*s on both standards of a 2-standard context, although the trends for the two classes of *O*s are in opposite directions.

4. The immediately preceding level of stimulation is a significant, though not the sole, contributory factor to the interaction effects shown by naive *O*s.

5. Interaction effects occur in a 2-standard context when standards are separated by as much as 1000 cycles. No systematic trend in the direction of a consistently increasing interaction effect is apparent, however, for increasing frequency differences between two standards over a difference range from 10 to 1000 cycles.

Several explanations, including those of 'central tendency' and 'assimilation,' of negatively directed constant error trends with increasing value of the standard have been discussed and evaluated. The reversal of this trend in the constant error for some of the practised *O*s has been explained on the basis of 'contrast.'

It is felt that there are several variables at work in determining the constant errors in judgments of stimuli distributed along some dimension and forming a context of judgment. For different *O*s these variables may differ in importance, resulting in individual differences in the sign, magnitude, and trend of the constant errors.

(Manuscript received November 26, 1945)

## References

1. Fernberger, S. W. Interdependence of judgments within the series for the method of constant stimuli. *J. exp. Psychol.*, 1920, 3, 126–150.

2. Freeman, G. L., & Sharp, H. Muscular action potentials and the time-error functions in lifted weight judgments. *J. exp. Psychol.*, 1941, 29, 23–36.

3. HOLLINGWORTH, H. L. The central tendency of judgment. *J. philos. psychol. sci. Meth.*, 1910, **7**, 461–469.
4. HOLLINGWORTH, H. L. The inaccuracy of movement. *Arch. Psychol.*, N. Y., 1909, No. 13.
5. KOESTER, T. The time error in pitch and loudness discrimination as a function of time interval and stimulus level. *Arch. Psychol.*, N. Y., 1945, No. 297.
6. LAUENSTEIN, O. Ansatz zu einer physiologischen Theorie des Vergleichs und der Zeitfehler. *Psychol. Forsch.*, 1933, **17**, 130–177.
7. MARTIN, L. J., & MUELLER, G. E. *Zur Analyse der Unterschiedsempfindlichkeit.* Leipsig: Barth, 1899, 233 pp.
8. NEEDHAM, J. G. The effect of the time-interval upon the time-error at different intensive levels. *J. exp. Psychol.*, 1935, **18**, 530–543.
9. STEVENS, S. S., & GERBRANDS, R. A twin-oscillator for auditory research. *Amer. J. Psychol.*, 1937, **49**, 113–115.
10. WOLFE, H. K. Untersuchungen über das Tongedächtniss. *Philos. Stud.* (Wundt), 1886, **3**, 534–571.
11. WOODROW, H. Weight-discrimination with a varying standard. *Amer. J. Psychol.*, 1933, **45**, 391–416.

# DEVELOPMENT OF SPEECH DURING INFANCY: CURVE OF PHONEMIC TYPES

BY ORVIS C. IRWIN

AND

HAN PIAO CHEN *

*State University of Iowa*

The literature on the mastery of speech sounds by infants has been summarized by the authors (3). The literature contains no clear-cut statements indicating the continual progress by the infant in the acquisition of the elemental sounds he will use when he becomes an adult. There are no systematic efforts to analyse the course of the development in early speech of the various classifications of speech sounds. The purposes of the present article are to determine the number of types of speech sounds appearing in infant vocalization during the first two and a half years of life, to indicate the nature of the development of speech sounds by means of a curve, and to derive its equation. Consideration is also given to sex differences. For these purposes, measures of central tendency are used. Problems of variability are being reserved for future publications.

In regard to the investigation itself, it was originally planned to visit a large number of babies in their homes twice each month during the first thirty months of life. Unfortunately the original plan was handicapped by the necessities of the war. Many infants moved with their families to other communities. Sickness, vacations, and other absences contributed the usual irregularities encountered in collecting the longitudinal type of materials. In spite of these handicaps, a very considerable amount of data has been collected and is available for analysis.

A total of 95 infants were included in the investigation. These infants were from middleclass homes, the parents being professional, business, clerical, and some laboring people. Although it has not been possible to follow all of them continuously throughout entire two and a half year period, the speech sounds of a majority of them were recorded for extended periods. About 20 babies have been followed to the end of the period of infancy and about 35 of them yield quite continuous records. The analysis consequently is in two parts, one of which includes the data obtained from the 95 cases, and a second which is limited to the most continuous data obtained from 35 cases. The data consists of 1622 records transcribed in the International Phonetic Alphabet. The record secured at a given visit contains a sample of such vocalizations as were uttered on 30 respirations or breaths (4). Observer and data reliabilities have been reported elsewhere (2).

The simplest index of phonemic development is furnished by the number of types of speech sounds. In this study the International Phonetic Alphabet as modified by Fairbanks (1) has been used. The time unit employed is the two-month period, so that the curve of development for the first 30 months is based upon 15 points, each point being a mean value for the corresponding two-month period. It has been found that the differences between means of adjacent months are not statistically different and that trends can be adequately shown with 15 points on the curve.

* Scholarship from the Committee on Wartime Planning for Chinese Students in the United States.

TABLE I

ALL DATA

| | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 | 12 | 13 | 14 | 15 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Means | 7.2 | 11.1 | 12.3 | 15.1 | 16.0 | 18.6 | 18.8 | 20.4 | 21.0 | 22.5 | 23.0 | 24.4 | 25.8 | 26.1 | 27.2 |
| *N* | 62 | 80 | 75 | 64 | 62 | 62 | 57 | 55 | 50 | 41 | 37 | 31 | 32 | 24 | 19 |
| Records | 125 | 181 | 161 | 170 | 147 | 149 | 127 | 128 | 106 | 84 | 68 | 52 | 51 | 41 | 27 |

MOST CONTINUOUS DATA

| | | | | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Means | 7.4 | 11.2 | 12.3 | 15.4 | 17.1 | 19.2 | 19.7 | 21.6 | 21.4 | 23.4 | 23.4 | 24.3 | 25.9 | 26.1 | 27.2 |
| *N* | 25 | 32 | 32 | 33 | 34 | 35 | 34 | 34 | 34 | 31 | 31 | 27 | 28 | 24 | 19 |
| Records | 52 | 74 | 71 | 94 | 87 | 91 | 81 | 85 | 74 | 70 | 62 | 48 | 47 | 41 | 27 |

Table I gives the means and the number of infants upon which these values are based, the number of records available, and also values for the 35 cases which include most of the continuous records. The table is to be read as follows: for the first period the mean number of phonemes is 7.2, the number of infants for this period is 62, and 125 transcriptions of 30 breaths each is the number of records.

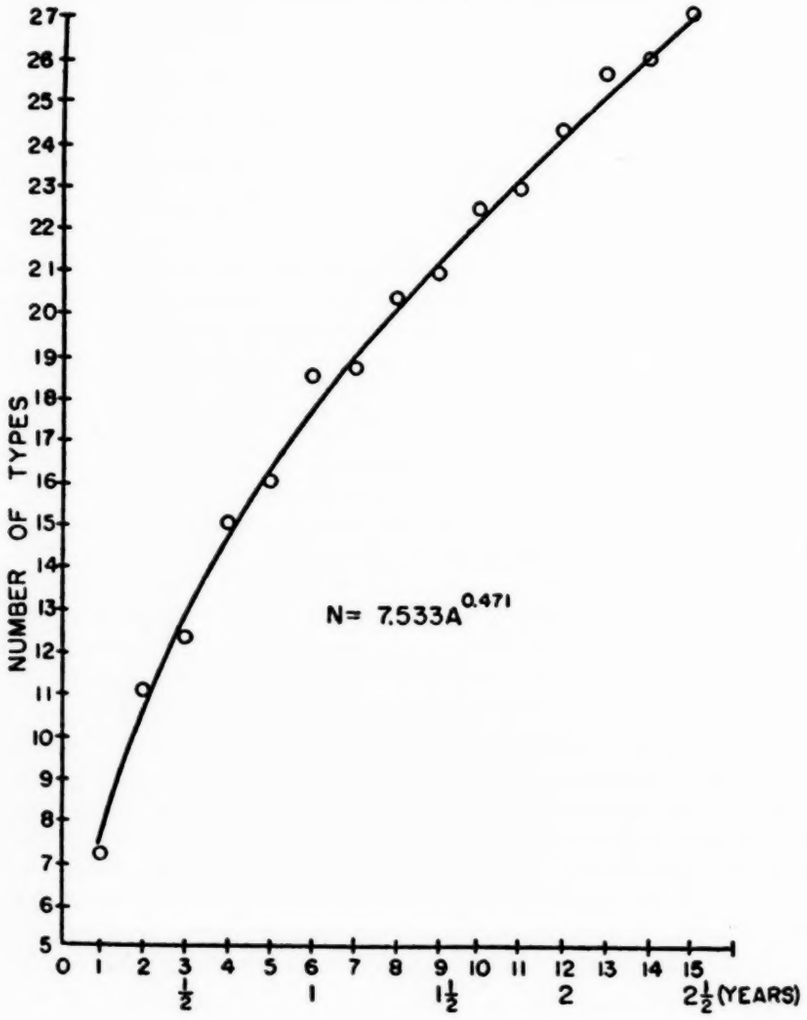


FIG. 1. Phoneme types (all data).

Fig. 1 presents the curve based on all the data made available by the investigation. Age ($A$) is laid off on the X axis, and ($N$), number of phonemic types, is plotted on the Y axis. Closed dots represent the observed values. The curve approximates a parabolic arc whose equation is $N = 7.533\ A^{0.471}$. By inspection it is obvious that the curve derived from the equation is a good fit to the empirical data. When the chi-square goodness of fit value is calculated it is found to be 0.191 ($P = .99$).

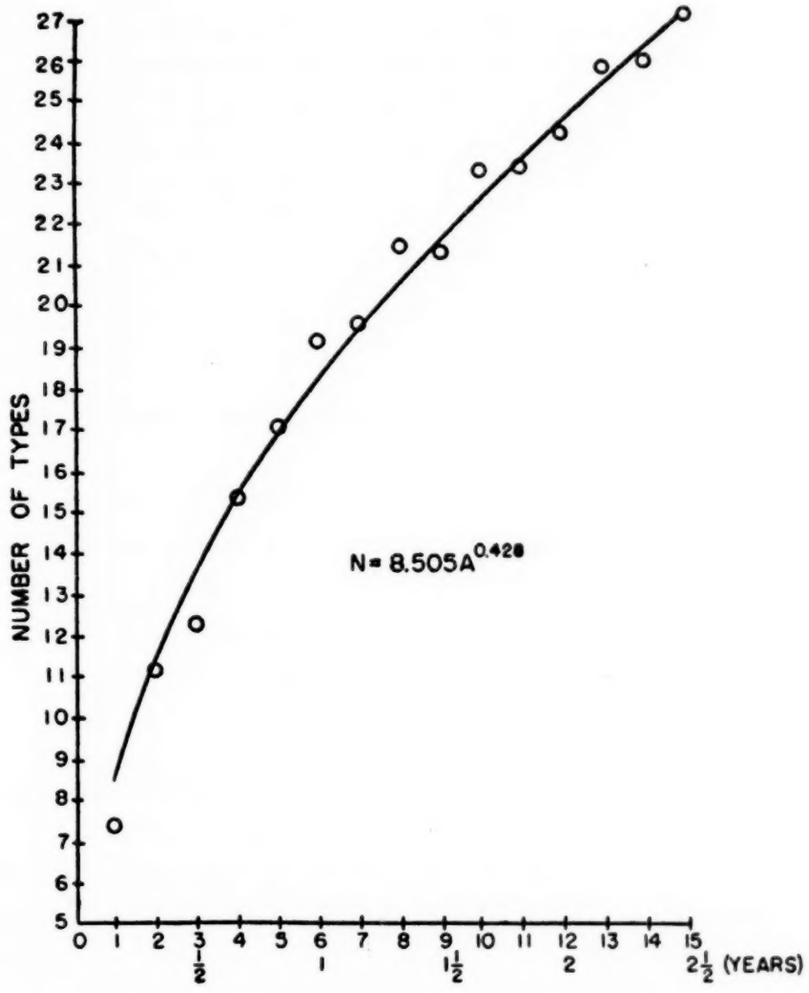


FIG. 2. Phoneme types (most continuous data).

In order to learn if those cases for whom there existed the most continuous records yield a reasonable approximation to the curve based on all data, Fig. 2 is presented. A total of 1004 records were available for this analysis. The equation for this curve is $N = 8.505\ A^{0.428}$ and likewise approximates a parabolic arc. This curve also is a good fit to the observed scores, the chi-square value being 0.422 ($P = .99$).

The two curves, it is seen, are quite similar and either may be taken as reasonably representative of the mastery of sound types

during the first two and one-half years of life. They show that the average number of types for the first two-month period is about eight and that for the last two-month period the infant is in possession of an average of 27 of the 35 sound elements used in adult English speech. It is interesting that the rate of the development of this type of infant behavior is a changing rather than a constant rate.

Table II gives the mean values for the fifteen units for boys and

TABLE II

DATA FOR BOYS AND GIRLS SEPARATELY

| | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 | 12 | 13 | 14 | 15 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Boys | 7.2 | 11.2 | 12.5 | 15.2 | 16.6 | 18.7 | 19.9 | 21.6 | 21.8 | 21.8 | 22.5 | 24.4 | 26.6 | 24.0 | 27.2 |
| *N* | 13 | 16 | 16 | 17 | 18 | 18 | 18 | 17 | 17 | 16 | 17 | 16 | 16 | 13 | 10 |
| Records | 27 | 35 | 37 | 53 | 48 | 52 | 47 | 46 | 42 | 37 | 35 | 29 | 32 | 20 | 15 |
| Girls | 7.7 | 11.1 | 12.1 | 15.5 | 17.6 | 20.1 | 19.2 | 21.6 | 20.8 | 25.0 | 24.4 | 24.1 | 24.8 | 28.5 | 27.1 |
| *N* | 12 | 16 | 16 | 16 | 16 | 17 | 16 | 17 | 17 | 15 | 14 | 11 | 12 | 11 | 9 |
| Records | 25 | 39 | 34 | 41 | 39 | 39 | 34 | 39 | 32 | 33 | 27 | 19 | 15 | 21 | 12 |

girls separately. These means are based upon the most continuous data.

When theoretical values are derived from the observed values in Table II and are plotted, the resulting curves are approximate parabolic arcs whose equations are:

$$\text{Boys, } N = 8.614\, A^{0.414}$$
$$\text{Girls, } N = 8.356\, A^{0.441}$$

The curves appear in Fig. 3.

The curves show that boys and girls start with the same endowment of speech sound types but that the girls end with a slightly higher achievement. They also show that the rate of development for the sexes changes. While the tendency for girls to be superior to boys is apparent, there are no statistically significant differences between the observed values. During the last year, where the tendency to differ is most apparent, of the six pairs of values only two are statistically significant. This is not enough evidence upon which to base an assertion of the definite presence of sex difference.

## SUMMARY

Ninety-five infants ranging in age from one to 30 months were visited in their homes and their spontaneously uttered speech sounds were recorded in the International Phonetic Alphabet. The sounds uttered on each of the 30 breaths constituted the sample obtained at each visit. The present article is concerned with the mean number

of phonemic types used by the infants and indicates the nature of this aspect of infant speech development. An equation is derived from the observed data and a curve is presented.

The curve approaches a parabolic arc, the equation being $N = 7.533\ A^{0.471}$. It shows that the phonemic equipment of the child during the first two-month period of life includes 7.5 sounds and that at the last two-month period (29–30 months) it is in possession of 27

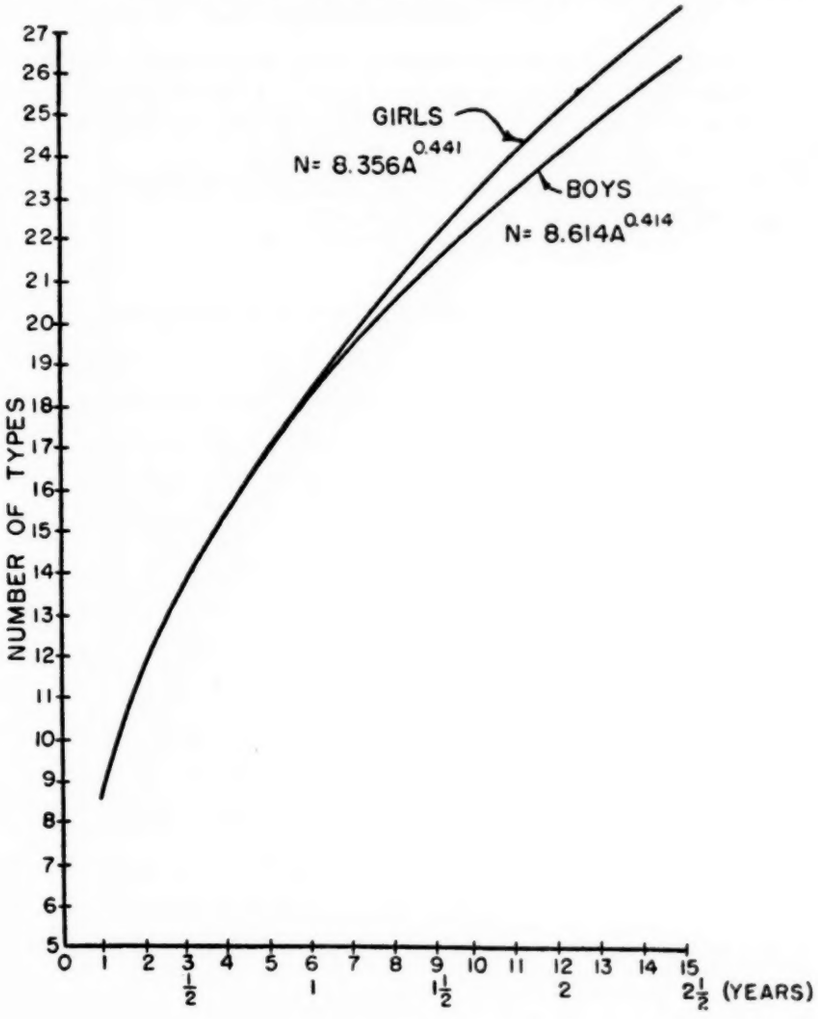


FIG. 3. Phoneme types (sex differences).

of the 35 sounds present in adult speech. The infant's mastery of the types of phonemes used in the English language proceeds in a curvilinear manner, that is, at a changing rate. Greater progress is made during the first year of life than during the second and a half years.

A second curve fitted to data transcribed from the vocalizations of the 35 children who were the most continuously observed infants among the group of 95, shows similar results. The equation for this curve is $N = 8.505\ A^{0.428}$. When equations of the mean values for boys and for girls are derived, the equations are $N = 8.614\ A^{0.414}$ for

boys and $N = 8.365\ A^{0.441}$ for girls. The curves derived from these equations show that while the two sexes begin with similar values and remain together during the first year, the girls tend to exceed the boys during and after the second year. The differences, however, are not statistically significant.

(Manuscript received December 26, 1945)

## References

1. Fairbanks, G. *Voice and articulation drillbook*, New York: Harper and Brothers, 1940.
2. Irwin, O. C. Reliability of infant speech sound data. *J. Speech Disorders*, 1945, **10**, 227–235.
3. Irwin, O. C., & Chen, H. P. Speech sound elements during the first year of life; A review of the literature. *J. Speech Disorders*, 1943, **8**, 121, 109.
4. Irwin, O. C., & Curry, T. Vowel elements in the crying vocalizations of infants under ten days of age. *Child Developm.*, 1941, **12**, 351–368.

# ADAPTATION IN ENERGY MOBILIZATION: CHANGES IN GENERAL LEVEL OF PALMAR SKIN CONDUCTANCE *

BY ELIZABETH DUFFY

*The Woman's College of the University of North Carolina*

AND

OLIVER L. LACEY

*The University of Alabama*

Palmar skin conductance (or the degree of conductivity shown by the skin of the palms of the hands when a minute electric current is passed through their surface) appears to be an indicator of the degree of energy mobilization of the individual at the moment the measurement is taken. The argument for this interpretation of palmar conductance has been ably presented by Darrow (**9**) and need not be repeated here. It may be pointed out, however, that what appears on the surface to be a wide diversity of interpretation among psychologists concerning the significance of skin conductance turns out upon closer examination to represent substantial agreement with respect to its meaning. Most of the studies of conductance have been concerned, not with the *general level* of conductance (which is the subject of the present study), but with the sharp *increase* in conductance (or the well known *galvanic skin response*) which occurs when certain stimulating agents are suddenly introduced. It is obvious, however, that the nature of the conditions producing an increase in conductance will throw light upon the significance of conductance in general. These conditions have been so diverse that they have tended to obscure the fact which appears to be of primary importance—namely, that *all the conditions under which an increase in conductance has been reported have in common an increase in energy mobilization.* Among such conditions we might list increased attention (**21**), startle or tension (**25**), sudden checks in the comprehension and solution of problems (**1**), heightened consciousness (**33**), conation (**2**),

* We wish to thank Dr. J. A. Highsmith, Head of the Psychology Department of The Woman's College, and the administration of the college, for an adjustment of teaching schedules which facilitated this investigation; Professor R. C. Davis, of Indiana University, for lending us the measuring device used in the study and for a critical reading of the manuscript; and Dr. Anna Reardon, of the Physics Department of The Woman's College, for advice and assistance in regard to apparatus.

states of anxiety (**32**), and anticipation, alertness, or apprehension (**8**). A survey of the literature to date affords ample justification, we believe, for interpreting palmar skin conductance as an index of energy mobilization.

The purpose of the present study was to investigate the *trends in energy mobilization* (as indicated by palmar conductance) *which occurred upon repetition, during a single experimental session and on successive days, of a certain task* to be described later, and to study also the *changes in conductance which occurred during the rest period preceding the task.* The question to be answered was *whether energy mobilization would show a 'learning curve' similar to that shown in overt behavior.* We wished to discover what adaptation, if any, would occur in palmar skin conductance.

Many investigators have demonstrated adaptation in the galvanic skin response in various situations. The response has been found to diminish or vanish under repeated visual (**27**), auditory (**6, 27**) and electrical (**23, 30**) stimulation; under unreinforced conditioned auditory stimulation (**16**); under repeated collapsing of a chair (**5**); and under reshowing of motion pictures (**28**). In most instances the phenomenon has been studied during a single experimental session only, but in the few cases where the investigation has extended over longer periods, adaptation has again been found (**28, 27, 30, 12**).[1] Indeed this finding has been called 'one of the most characteristic phenomena of the galvanic skin reflex'—though one infrequently investigated (**6**).

This generally observed diminution of the galvanic skin response, however, is only indirectly relevant to our problem. We are concerned directly, not with the adaptation of the *galvanic response* to a particular stimulus, but with the adaptation of the *general level* of conductance measured over a considerable period of time. Such adaptation to repeated application of thermal stimulation has been reported by Maslov (**26**), and to continuous application of auditory stimulation by Davis (**11**). Seward and Seward (**30**) have reported a *decline* of resistance during an eight minute period in which five electric shocks were administered. No consistent upward or downward trend of resistance from day to day was found.

The study most closely related to our own, however, is a second investigation by R. C. Davis (**12**), which deals, at least in part, with the general level of conductance, although *change* in conductance is the major concern of the paper. In a series of seven experiments, Davis studied 'the habituation of the galvanic reflex and body resistance to

[1] Farmer and Chambers (**17**), however, made the incidental observation that the adaptation was only temporary, since the response reappeared with the same stimulus after the subject had spent a week-end in the country.

repetitions of various types of stimuli' (**12**, p. 532). The number of *Ss* in the experiments ranged from 5 to 10, and the repetition of stimuli occurred in most instances on five consecutive days, though in one experiment the *Ss* were re-tested on the 6th and the 33rd days after the end of the five-day period. As we have mentioned, Davis was primarily interested in the *change* in resistance during a work period or other situation rather than in the *average* resistance during the period. Hence his tables and graphs deal chiefly with the former phenomenon, which is of only incidental interest for our purposes. Two of his conclusions, however, pertain directly to the question of general level: 1. "Initial resistance regularly showed a pronounced increase over a period of five days"; 2. In the group of *Ss* re-tested six and 33 days after the end of the original series, "the initial resistance tended to remain at the same level as on the last day of the habituation series" (**12**, p. 533).

Our own investigation was designed to study some of the same problems investigated by Davis. It differs from Davis's study in the following respects: (1) the primary concern was with *average* resistance rather than with *changes* in resistance; (2) palmar, rather than finger-tip, resistance was recorded; (3) the stimulus situation was of a different kind; (4) the stimulus situation was repeated four times during each experimental session; (5) the statistical treatment of the data, in harmony with recent practice in physiological studies, is more complete; (6) resistance was translated into its reciprocal—conductance.

## Description of the Experiment

Palmar skin resistance was recorded for 10 women students during a period when their auditory limens were being obtained, and also during a rest period preceding the liminal determinations. The procedure for determining the limen was repeated four times at each experimental session, with an interval of two min. between successive trials. The entire procedure was repeated for each *S* on three consecutive days at the same hour.

### Apparatus

1. *Apparatus for measurement of skin resistance*

The instrument used for measurement of skin resistance was very kindly lent to us by R. C. Davis of the Department of Psychology of Indiana University. An excellent description of this device is to be found in an article by Davis and Porter (**10**). As this article points out, the instrument has the advantage of passing through the *S* a current which is essentially constant. In the study reported here the current was obtained from storage batteries and was maintained at about 80 microamperes. With the Davis instrument the deflections of the milliameter are proportional to the dermal response (**20**).

Contact was made with the *S* by means of rubber cup electrodes attached to the palms. These electrodes were constructed from common bathtub stoppers. They had a diameter of 2.5 cm. and a depth of .6 cm. The bottom of the stopper was covered with a zinc plate. A zinc post, passed through the stopper and soldered to the plate, made contact with leads from the

galvanometer. The cups were filled with an electrode jelly composed of 1 percent Bacto-Agar, 1 percent zinc sulphate.

2. *Apparatus for determination of auditory limens*

The auditory limens were measured by use of a Western Electric 6B audiometer. In order to obtain a finer variation in the output a Daven attenuator was inserted in the lines to the earphones. This combination permitted variation of auditory output in steps of one-half decibel.

## Method

After the *S* had been seated for approximately 11 min., during which the electrodes were attached and the earphones were adjusted, she was put in circuit with the measuring device and given a rest period of 5½ min. During the last 2½ min. of this rest period (after the current had been flowing through the circuit for three min.) readings of resistance were taken at 15-sec. intervals. The readings taken during this period will be reported under the heading 'Rest.' At the end of this period a ready signal was given, and 30 sec. later the first auditory stimulus was presented.

The procedure for the liminal determinations was as follows: A light of two sec. duration warned the *S* to listen for an auditory stimulus. One second after the light was extinguished the *E* presented (or did not present, according to a previously determined random arrangement) an auditory stimulus of zero db. (Each stimulus given lasted for a period of two sec.) If the *S* heard a tone, she tapped lightly on the floor with her foot; if she heard no tone, she made no response.

Ten sec. after the first light signal the procedure was repeated. Upon this second presentation, a tone was given if no tone had been given upon the first presentation. Conversely, no tone was given upon the second signal if a tone had been given on the first.

The stimulus intensity was now reduced 2½ db. by means of the audiometer, and the entire procedure was repeated. This process was continued (the strength of the auditory signal being reduced in steps of 2½ db.) until the *S* failed to discriminate correctly between tone and no-tone. (The *S* was considered to fail whenever she did not respond to the tone, *or* she reported a tone when no tone was given.)

After the limen was thus roughly estimated the audiometer was set to the last 5-db. step at which the *S* had discriminated correctly. With the audiometer thus set, the limen was determined more accurately by continuing in an entirely similar manner, using, however, the half-db. steps provided by the attenuator. The series ended when the *S* failed to discriminate for two successive stimulus levels.

The entire procedure took ordinarily between four and nine min. During the two min. between successive series the *S* remained seated but was permitted to shift position and to talk.

Resistance readings were taken at 15-sec. intervals throughout the period when the limen was being determined. It was impossible to maintain a constant time period for this procedure, since the time required to obtain a limen cannot be controlled. In studying adaptation effects the average resistance level was computed for two different parts of this period, which we shall call Task I and Task II. Task I consisted of the first 2½ min. of the period during which the limen was being obtained. Here the auditory discriminations were quite easy. Task II consisted of the last 1½ min. of the period during which the limen was being obtained. The discriminations here were more difficult since they were close to the *S*'s limen.

It should be pointed out parenthetically that the procedure of recording resistance levels regularly at 15-sec. intervals, rather than recording a series of minima, differentiates our study from studies of the galvanic skin response and places it in the category of investigations of the level of skin resistance. Since a galvanic response consists of a decrease in resistance, the occurrence of one would, of course, have the effect of lowering the next resistance reading. However, in our investigation, resistance levels were recorded at regular time intervals and not merely when a galvanic response occurred. As Davis points out (12, p. 507), this procedure has the advantage of making possible "a statement of the average conductance of a group of subjects at a series of given moments, whereas knowledge of a series of minima scattered irregularly in time, which the ordinary technique of recording deflections yields, would provide no means whatever of grouping data unless approximate temporal coincidence of deflections in different subjects can be insured."

All resistance measures were translated into the reciprocal value—conductance. This

procedure was followed in view of the evidence (7) that the amount of palmar perspiration tends to vary as the reciprocal of the resistance change.[2]

Temperature and humidity were recorded at the end of each experimental session.

Before the session began, the *S* washed her hands in lukewarm water without soap. Care was taken to avoid variations in the contact and pressure of the electrodes, and to avoid movements of the *S* during the period when resistance was being measured.

## Discussion

Fig. 1 summarizes the major findings. *Conductance decreases:* (1) *from series to series on a given day, and* (2) *from day to day.* The

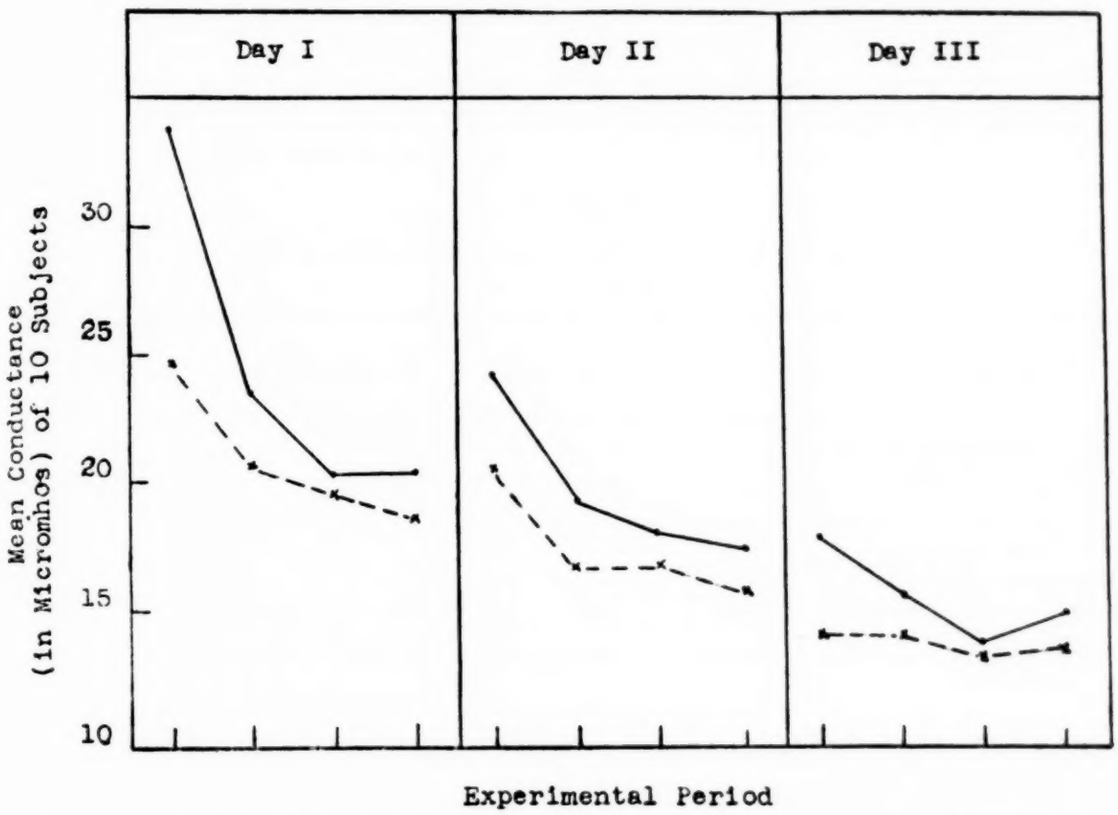


Fig. 1. Palmar conductance during easy (Task I) and difficult (Task II) discriminations. (Task I ———; Task II – – –)

first series on a succeeding day usually begins with conductance which is *slightly greater* than that of the *last* series on the preceding day, but *considerably less* than that of the *first* series on the preceding day. In other words, some degree of recovery of responsiveness to the situation occurs between experimental sessions, even when those sessions

[2] The use of conductance as a measure of the galvanic properties of the skin has recently been questioned by Haggard (22), who presents evidence indicating that a modification of log. resistance is preferable for measuring the GSR and who suggests further that this same unit is appropriate also for comparison of general levels of skin conductance. The junior author, however, has recently completed a methodological study (24) of this question (now in preparation for publication) which indicates that conductance as such is an appropriate measure of these electrical properties of the skin. Other units of course (such as that proposed by Haggard) may perhaps also be suitable.

come, as in the present instance, on successive days. Yet the downward trend of conductance throughout the course of the experiment is quite marked. A more detailed analysis of the findings is presented below.

*Within-the-day adaptation*

Adaptation of skin conductance in the course of the four repetitions of the stimulus situation which occurred during each experimental session is shown by analysis of two sets of data:

(1) The mean conductance of the 10 *Ss* for each of the four repetitions of Task I (Table I).

(2) The mean conductance of the 10 subjects for each of the four repetitions of Task II (Table I).

TABLE I

ADAPTATION WITHIN EACH DAY SHOWN BY THE FOUR PERIODS OF TASKS I AND II

| Period | 1 | 2 | 3 | 4 | Probability of Chance Occurrence |
|---|---|---|---|---|---|
| Mean Conductance in Micromhos During Task I..... | 25.16 | 19.52 | 17.38 | 17.57 | $P < .01$* |
| Mean Conductance in Micromhos During Task II..... | 19.72 | 17.08 | 16.54 | 15.82 | $P < .01$* |

* Significant.

Table I clearly demonstrates that skin conductance decreases with repetition of the task during a single experimental session. This is true for both the easy auditory discriminations (Task I) and the more difficult auditory discriminations (Task II). A continued decrease in conductance is somewhat more apparent for the more difficult task (Task II). The fact that the level of conductance is lower for Task II than for Task I should be interpreted in the light of the fact that the four periods of Task II occurred in each instance later in a given series than the corresponding periods of Task I.

That the curve of adaptation may be logarithmic in nature is suggested by the fact that Periods 3 and 4 are in both cases closer in conductance values than are Periods 1 and 2. (Indeed, Period 4 shows a slight rise in conductance during Task I.) However, the data are clearly insufficient for more than suggesting such a relation. We can say with confidence that adaptation occurs, but the nature of the curve of adaptation we can only surmise.

*Day-to-day adaptation*

Adaptation of conductance from day to day during the three-day period of the experiment is shown by analysis of three sets of data:

(1) The mean conductance of the 10 Ss during the rest period, or fore period, on each of the three days.

(2) The mean conductance during the first performance of Task I on each of the three days. (Only the *first* of the four performances each day was used in the analysis, since the time interval for the other three could not be rigidly controlled.)

(3) The mean conductance of (1) and (2) above combined.

TABLE II

ADAPTATION FROM DAY TO DAY AS SHOWN BY DAILY CONDUCTANCE DURING REST PERIOD, TASK I, AND REST PERIOD AND TASK I COMBINED

| | Day I | Day II | Day III | Probability of Chance Occurrence |
|---|---|---|---|---|
| Mean Conductance in Micromhos During Rest Period..... | 25.68 | 23.24 | 15.88 | $.05 < P < .10$* |
| Mean Conductance in Micromhos During Task I..... | 33.79 | 23.97 | 17.74 | $P < .01$† |
| Mean Conductance in Micromhos During Rest Period and Task I Combined..... | 29.73 | 23.60 | 16.81 | $P < .01$† |

* This probability, as estimated by interpolation, is about .06—just above the conventional 5 percent level of significance.

† Significant.

The first line of entries in Table II shows the trends in conductance during the rest period preceding the auditory series on three consecutive days. It will be observed that there is decreasing conductance from Day I to Day III and that the difference in mean conductance, which analysis of variance shows to have a probability of .06, is just below the conventional level of significance. Davis (12) reports finding no adaptation occurring during the rest period, or fore period, in the course of the five days of his experiment.

The second line of entries in Table II shows the mean conductance for the first performance of the easy auditory discriminations, or Task I, on each of the three days. There is decreasing conductance from Day I to Day III, and this finding is highly significant statistically.

The third line of entries in Table II shows the mean conductance on each of the three days for the rest period and the first performance

of Task I when these two periods are combined. Again there is decreasing conductance from Day I to Day III, and again the finding is highly significant statistically.

Table II shows adaptation in conductance from Day I to Day III; it does not show the significance of the difference in conductance between pairs of days. It is of interest to know not merely that there is adaptation in conductance over a three-day period, but also whether there is a significant difference between Day I and Day II, and whether the adaptation effect continues so that Day II is significantly different from Day III. Table III presents this analysis for the rest period, for the first performance of Task I, and for the rest period and the first performance of Task I combined.

The first line of values in Table III shows that, for the rest period,

TABLE III

AVERAGE DIFFERENCE BETWEEN DAYS IN CONDUCTANCE

| Mean Conductance During: | Day I–Day II | | Day II–Day III | | Day I–Day III | |
|---|---|---|---|---|---|---|
| | Mean Difference | Probability | Mean Difference | Probability | Mean Difference | Probability |
| Rest Period..... | 2.44 | P>.10 | 7.35 | .02<P<.05* | 9.79 | P<.01* |
| Task I..... | 9.82 | .01<P<.02* | 6.23 | .02<P<.05* | 16.04 | P<.01* |
| Rest Period and Task I Combined | 6.13 | .02<P<.05* | 6.79 | .02<P<.05* | 12.92 | P<.01* |

* Significant.

Day I is *not* significantly different from Day II, Day II *is* significantly different from Day III, and Day I *is* significantly different from Day III. These results should be compared with those found above by analysis of variance. It will be recalled that the earlier analysis indicated that the differences from day to day just fail to reach the conventional level of significance. The present technique of analysis (the *t*-test), however, would appear to establish reliably the occurrence of adaptation over the three-day period and from Day II to Day III. The failure of the variance analysis to demonstrate this effect is clearly due to the fact that Day I and Day II do not differ significantly.

The second line of entries in Table III shows that, for Task I, each day is significantly different from every other day, or that the adaptation effect is significantly shown from day to day. The *S* does not adjust to the experimental situation in the first day and then show no further decrease in conductance. On the contrary, adaptation continues from Day II to Day III.

The tables thus far presented show conclusively that there is a relationship between conductance and day. Unfortunately, we must at this point raise the question whether this is a true adaptation effect or whether the relationship with day is really a relationship with the *temperature* of the day. Table IV shows that temperature declined

TABLE IV

AVERAGE OF DAILY ATMOSPHERIC TEMPERATURES FOR THE 10 SUBJECTS

| | Day I | Day II | Day III | Probability of Chance Occurrence |
|---|---|---|---|---|
| Average of Temperatures for Day..... | 29.84° C. | 29.07° C. | 27.77° C. | $P < .01$* |

* Significant.

during the course of the experiment, and that temperature and day are significantly connected. (Humidity was not found to be significantly related to day and may, therefore, be disregarded.)

Since, as shown by Table IV, the previously described relations of conductance to day may be due either to adaptation or to temperature, further analysis is required. Partial correlation methods are not possible since measurements on only three days are available for each subject. An attempt to use this form of analysis would leave zero degrees of freedom, and thus no test of significance could be applied. The following technique of analysis was therefore used:

A. (1) A regression line was calculated for each subject, relating temperature to conductance.

(2) Deviations from regression were analyzed for significant dependence upon day.

B. (1) A regression line of conductance upon day was calculated for each subject.

(2) Deviations from regression were analyzed for significant dependence upon temperature.

The rationale is as follows: If deviations from temperature regression are significantly related to day, then adaptation is shown. If, on the other hand, deviations from day regression are significantly related to temperature, then a temperature effect is shown.

The results of these analyses are shown in Tables V, VI, and VII below. The analysis was in each case performed for the conductance during the rest period combined with the conductance during the first performance of Task I on each of the three days.

Table V shows that the *deviations of conductance from temperature regression are significantly related to day.*

TABLE V

AVERAGE DEVIATION OF THE CONDUCTANCE (DURING THE REST PERIOD AND TASK I COMBINED) FROM THE EXPECTED VALUE (CALCULATED FROM TEMPERATURE)

| | Day I | Day II | Day III | Probability of Chance Occurrence |
|---|---|---|---|---|
| Mean Deviation from Regression (in Micromhos)..... | 7.48 | 0.54 | −8.10 | $P < .01$* |

* Significant, provided the test is valid. For a discussion of the validity of the test, see the text.

TABLE VI

THE SIGNIFICANCE OF THE DIFFERENCE BETWEEN DAYS IN AVERAGE DEVIATION OF THE CONDUCTANCE (DURING THE REST PERIOD AND TASK I COMBINED) FROM THE EXPECTED VALUE (CALCULATED FROM TEMPERATURE)

| Days | I vs. II | II vs. III | I vs. III |
|---|---|---|---|
| P | $.05 < P < .10$ | $.05 < P < .10$ | $P < .01$* |

* Significant.

TABLE VII

AVERAGE DEVIATION OF THE CONDUCTANCE (DURING THE REST PERIOD AND TASK I COMBINED) FROM THE EXPECTED VALUE (CALCULATED FROM DAY)

| Temperature | Highest | Middle | Lowest | P |
|---|---|---|---|---|
| Deviation in Micromhos | −.22 | −3.03 | 3.27 | $P > .10$ |

Table VI shows the significance of the difference between pairs of days. The difference between Day I and Day II closely approaches the five percent level of significance; the difference between Day II and Day III also closely approaches the five percent level; and Day I differs from Day III at the one percent level of significance. Thus adaptation is seen to be significant over the three-day interval and is not to be attributed to adaptation between Day I and Day II alone, nor to adaptation between Day II and Day III alone.

To this point we have shown adaptation effects to be significant with or without correction for tempeature. Table VII answers the converse question: Is the temperature effect significant if the readings are corrected for adaptation effect? Temperature effect is *not* found significant if the day effect is ruled out, since the probability of chance occurrence is greater than .05 (its actual value is about .20). It thus appears that the true picture of day-to-day adaptation is shown in Tables II and III and that (for our data) temperature need not be considered at all.

A question remains, however, as to the validity of the tests of significance which we have applied to the data corrected for regression. The techniques used (the *t*-test and analysis of variance) would assume that these data follow the normal curve. It is known, however, that if the raw data are normally distributed then data adjusted for regression in the way described above will *not* be so distributed. Conventional tests of significance are therefore not strictly valid. It is theoretically possible to devise an exact test of significance for data corrected in this fashion, but the labor involved is prohibitive.[3] (Parenthetically it may be remarked that an analogous situation exists, for example, in the fitting of logarithmic regression lines.) In justification of our use of the *t*-test we may say: (1) that even though an underlying distribution be not normal, the distribution of means of samples from that distribution will tend toward normality, and it is means that we are actually comparing; and (2) that, since the tests used indicate differences significant at the one percent level, we may reasonably claim adaptation to be shown at the five percent level.

As further argument for disregarding the effect of temperature, we may say that, not only did we ourselves find that temperature was not significantly related to conductance when the effect of day was ruled out, but a number of other investigators of palmar resistance have also discounted the effect of temperature. Freeman and Giffin (**18**, p. 67), for example, state that "changes in room temperature or humidity have little influence on resistance measures. . . . If readings had been secured from non-palmar areas greater effects of temperature and humidity would probably have been shown, as sweating here is largely thermo-regulatory." Darrow and Freeman (**8**, p. 747) refer to "the finding of Kuno and Ikeuchi that increasing atmospheric temperature caused increased non-palmar, non-plantar sweating, while there was no increase in perspiration on the palms." Smith (**31**) states that Richter found a *decrease* in resistance with an increase in temperature; that Cattell found a small *increase* in resistance in hot weather; and that he himself found no consistent effect of temperature or humidity on apparent resistance level.

Our data provided two additional lines of evidence which appear also to support the belief that temperature may be disregarded: (1) In the case of two *S*s the temperature *rose* over the interval from Day I to Day III. These *S*s, nonetheless, show the same progressive

[3] We wish to thank Professor Mark Kac, Department of Mathematics, Cornell University, for his advice upon this point. In a personal communication he states that "although strict mathematical justification is lacking, the procedure seems appropriate on 'intuitive' grounds." It is further worth mention that the data corrected for temperature effects in this way do not deviate significantly from normality in so far as symmetry is concerned.

*decline* in conductance as do the others. (2) The correlations of temperature with average *S* conductance for each day considered separately are not significant (Day I, $r = .035$; Day II, $r = .189$; Day III, $r = .344$).

In conclusion, we may say that it seems probable that Tables II and III present a true picture of adaptation in conductance, and that these tables show a significant decrease in conductance from day to day.

*Adaptation during the rest period, or fore period*

It has been said that a falling off in average resistance (i.e., an increase in conductance) during the fore period, when the individual is subjected to no controlled stimulation, is the regular finding (**12**). This has frequently been attributed to physical factors. Davis, however, found the usual decrement in resistance during a period in which the *S*s were in no way connected to the instrument. He, therefore, concluded that "a substantial part of this 'adaptation curve' is not an artifact of measurement, but a response to the experimental situation" (**12**, p. 533).

Our own records show *increasing*, rather than decreasing, resistance during this period. We have 30 records of resistance trends during the fore period, or rest period (three records for each of 10 *S*s). Of these, 16 records clearly show increases in resistance, 10 are somewhat irregular but show a tendency toward an increase, while only three records show a clear *decrease*, and one record, though irregular, shows a tendency toward a decrease. In line with our findings are those of Bagchi and Greenwald (**3**) who report that the control period (five min.) showed a slow decrease in conductance (or increase in resistance). They state that this decrease was also noted by White (**34**). It thus appears that adaptation may occur even during a period of relatively uncontrolled stimulation, on any given day and, as we have pointed out above, from day to day (see Table II).

## Summary and Conclusions

In a study of adaptation in energy mobilization (as indicated by a decreasing level of palmar conductance), resistance measurements were obtained during successive rest periods and successive work periods.

A series of tones, varying in intensity from zero db. to a point just below the *S*'s limen, was presented at regular intervals to 10 *S*s, four times during an experimental session, on three consecutive days at the same hour. Whenever the *S* heard a tone, she indicated it by a light tap of the foot.

Palmar resistance was recorded at 15-sec. intervals, both throughout the stimulus series and during the rest period preceding the series on each day. Resistance measures were translated into the reciprocal value known as conductance and were analyzed for adaptation effects.

*Decreasing conductance*, indicating a *decrease in energy mobilization*, was found:

1. In the course of a $2\frac{1}{2}$-min. rest period preceding the presentation of stimuli.

2. During the four repetitions of auditory discriminations which occurred in a single experimental session.

3. From day to day throughout the three-day period of the experiment, during both (1) the rest period, and (2) auditory discriminations.

*Adaptation effects in energy mobilization are thus shown to occur upon the repetition of a stimulus situation within a given experimental session and from day to day.* Progressively less energy is mobilized in adjusting to the general experimental situation (as shown in the decrease of conductance during the rest period from day to day). Progressively less energy is mobilized for carrying out a task (as shown in the decrease of conductance from series to series of auditory discriminations on the same day and on successive days).

That useless overt movements tend to disappear in the course of 'learning,' or during successive responses to a given stimulus situation, is a well recognized fact. It is usually said also that, in the course of 'learning,' a task is performed with a smaller expenditure of energy, though energy expenditure upon repetition of a task has been infrequently measured. Becker and Olsen (**4**), measuring the effect of practice upon the metabolic 'cost' of learning nonsense syllables, have reported that practice increases the number of lists learned in a given time and reduces the cost per unit of time. Rounds, Schubert and Poffenberger (**29**), measuring the metabolic change during mental work, found that it remained practically constant at successive sittings, but, since the number of problems solved increased, there was a considerable decrease in the expenditure of energy per problem. On the other hand, Davis (**12**, p. 522), measuring skin resistance during work on problems in addition, reported "no discernible decrease in the resistance change accompanying mental work when the performance is repeated." His tables, moreover, show no appreciable change in general level of resistance. It is difficult to account for this finding (since here there was no increase in the number of problems solved) except, perhaps, in the light of the shortness of the practice period. (His *S*s worked for two min. on five days.) In our own experimental situation the performance (auditory discriminations at regular intervals) was of such a kind that there was no

opportunity for an increase in the number of units of work during a given time. Under these conditions, we find a very definite decrease in the energy mobilization during the task. The nature of the task was such as to rule out the possibility that this decrease was due merely to a decrease in overt movements. It was undoubtedly due to some change occurring in implicit processes. A part of the change could no doubt be explained in terms of general adjustment or adaptation to the experimental situation as a whole, since adaptation occurred during the rest period. Since, however, the initiation of the task always caused an increase in conductance followed by a more or less continuous decrease during the course of the task, it seems probable that some part of the adaptation effect is to be attributed to adaptation to the specific features of the task itself.

We have, then, we believe, in this demonstration of decrease in energy mobilization during repetition of a task, a phenomenon which is similar to the well recognized decrease in useless overt movements sometimes represented in the form of a 'learning curve.' The internal processes of energy mobilization, like the overt responses of the individual, show increasing adaptation to the situation or, if you like, 'learning.' It seems likely, on theoretical grounds as well as on the basis of the fragmentary evidence found in our data, that the curve of adaptation for energy mobilization will, like the typical learning curve, be found to be logarithmic in nature.

Placed in a broader setting, our findings emphasize the organismic principle of response of the organism as a whole. Implicit and overt activities are merely separate aspects of the total adjustment of the entire organism; there are 'internal' as well as 'external' evidences of 'learning.' Implicit and overt activities vary together as discriminable aspects of a unitary phenomenon of adjustment. Both implicit and overt behavior become ever more finely adjusted to the requirements of a situation.

Though we may be accused of going afield to drag in an old argument, we cannot resist pointing out that the nicety of adjustment of energy mobilization to the requirements of the situation seems to fit in better with a theory of a continuum of degrees of energy mobilization (**13, 14, 15**) than with the more generally accepted theory of a dichotomy between 'normal' or 'non-emotional' energy mobilization on the one hand and the energy mobilization of 'emotion' on the other. It would be a strained interpretation, indeed, to assume that the continuous decrease in energy mobilization shown in our study is to be accounted for solely in terms of a decrease from series to series and from day to day in the so-called 'emotional' factors in the situation. According to this interpretation, we suppose it would be necessary to assume that, with continued repetitions, the emotional

factors would finally be eliminated and a constant 'non-emotional' energy level be maintained.

But leaving aside theoretical controversy, we believe that the significance of this study lies in the fact that we have, perhaps for the first time, applied to measures of level of palmar conductance during repetitive stimulation a statistical analysis which shows that decreases in conductance from series to series and from day to day satisfy conventional statistical requirements for significance and cannot be considered a mere chance variation. Thus we have shown that energy mobilization (as indicated by palmar conductance) follows a definite course of adaptation.

(Manuscript received January 3, 1946)

## References


1. Abel, T. M. Attitudes and the galvanic skin reflex. *J. exp. Psychol.*, 1930, **13**, 47–60.
2. Aveling, F. The conative indications of the psychogalvanic phenomena. *Proc. Eighth Int. Cong. Psychol.*, 1926, **8**, 227–234.
3. Bagchi, B. K., & Greenwald, D. U. Electrodermal changes during certain types of attention. *J. gen. Psychol.*, 1937, **17**, 235–248.
4. Becker, F. C., & Olsen, H. Metabolism during mental work. *Skand. Arch. Physiol.*, 1914, **31**, 81–197.
5. Blatz, W. E. The cardiac, respiratory, and electrical phenomena involved in the emotion of fear. *J. exp. Psychol.*, 1925, **8**, 109–132.
6. Coombs, C. H. Adaptation of the galvanic response to auditory stimuli. *J. exp. Psychol.*, 1938, **22**, 244–268.
7. Darrow, C. W. The significance of skin resistance in the light of its relation to the amount of perspiration. *J. gen. Psychol.*, 1934, **11**, 451–452.
8. Darrow, C. W., & Freeman, G. L. Palmar skin-resistance changes contrasted with non-palmar changes and rate of insensible weight loss. *J. exp. Psychol.*, 1934, **17**, 739–748.
9. Darrow, C. W. The galvanic skin reflex (sweating) and blood-pressure as preparatory and facilitative functions. *Psychol. Bull.*, 1936, **33**, 73–94.
10. Davis, R. C., & Porter, J. M., Jr. A measuring device for the galvanic reflex. *J. gen. Psychol.*, 1931, **5**, 115–120.
11. Davis, R. C. Electrical skin resistance before, during, and after a period of noise stimulation. *J. exp. Psychol.*, 1932, **15**, 108–117.
12. Davis, R. C. Modification of the galvanic reflex by daily repetition of a stimulus. *J. exp. Psychol.*, 1934, **17**, 504–535.
13. Duffy, E. Emotion: an example of the need for reorientation in psychology. *Psychol. Rev.*, 1934, **41**, 184–198.
14. Duffy, E. The conceptual categories of psychology: a suggestion for revision. *Psychol. Rev.*, 1941, **48**, 177–203.
15. Duffy, E. An explanation of 'emotional' phenomena without the use of the concept 'emotion.' *J. gen. Psychol.*, 1941, **25**, 283–293.
16. Ellson, D. G. Spontaneous recovery of the galvanic skin response as a function of the recovery interval. *J. exp. Psychol.*, 1939, **25**, 586–600.
17. Farmer, E., & Chambers, E. G. Concerning the use of the psycho-galvanic reflex in psychological experiments. *Brit. J. Psychol.*, 1924–25, **15**, 237–254.
18. Freeman, G. L., & Giffin, L. L. The measurement of general reactivity under basal conditions. *J. gen. Psychol.*, 1939, **21**, 63–72.
19. Freeman, G. L., & Katzoff, E. T. Methodological evaluation of the galvanic skin response, with special reference to the formula for R. Q. (recovery quotient). *J. exp. Psychol.*, 1942, **31**, 239–248.

20. GREENWALD, D. U. Circuits now available for the measurement of electrodermal responses. *Psychol. Bull.*, 1935, **32**, 779–791.
21. GRÜNBAUM, A. A. Aufmerksamkeit, Emotivität und galvanisches Phänomen bei morbus Basedowi. *Psychiat. neurol. Bl.*, 1922, **26**, 359–371.
22. HAGGARD, E. A. Experimental studies in affective processes: II. On the quantification and evaluation of 'measured' changes in skin resistance. *J. exp. Psychol.*, 1945, **35**, 46–56.
23. HOVLAND, C. T., & RIESEN, A. H. Magnitude of galvanic and vasomotor response as a function of stimulus intensity. *J. gen. Psychol.*, 1940, **23**, 103–121.
24. LACEY, O. L. The units of measurement of the level of galvanic skin characteristics (In preparation).
25. LANDIS, C., & HUNT, W. A. The conscious correlates of the galvanic skin response. *J. exp. Psychol.*, 1935, **18**, 505–529.
26. MASLOV, A. The effects of local application of cold upon the electrical resistance of the skin. *Fiziol. Zh. S. S. S. R.*, 1940, **28**, 264–270.
27. PORTER, J. M., JR. Adaptation of the galvanic skin response. *J. exp. Psychol.*, 1938, **23**, 553–557.
28. RASMUS, M. H. Degeneration of emotional response upon reshowing of motion picture situations. *Psychol. Monog.*, 1936, **48**, 40–56.
29. ROUNDS, G., SCHUBERT, H., & POFFENBERGER, A. T. Effects of practice upon the metabolic cost of mental work. *J. gen. Psychol.*, 1932, **7**, 65–79.
30. SEWARD, J. P., JR., & SEWARD, G. H. The relation of galvanic skin reactions to preceding resistance. *J. exp. Psychol.*, 1935, **18**, 64–79.
31. SMITH, C. E. The effect of changes in the temperature and humidity of the air on the apparent skin resistance. *J. Psychol.*, 1937, **3**, 325–331.
32. SOLOMON, A. P., & FENTRESS, T. L. Galvanic skin reflex and blood pressure reactions in the psychoneuroses. *J. nerv. ment. Dis.*, 1934, **80**, 163–182.
33. WEINBERG, A. A. Psyche und unwillkürliches Nervensystem. *Z. ges. Neurol. Psychiat.*, 1923, **85**, 543–565; **86**, 375–390; 1924, **93**, 421–445.
34. WHITE, M. M. Relation of bodily tension to electrical resistance. *J. exp. Psychol.*, 1930, **13**, 267–277.

# VALIDITY OF FREQUENCY OF BLINKING AS A CRITERION OF READABILITY *

BY MILES A. TINKER

*University of Minnesota*

Readability refers to the ease and efficiency with which material is read. It may be influenced by different conditions, including variation in such typographical factors as size of type, type form, and style of type face. In general, those factors which facilitate the speed or efficiency with which printed material can be assimilated are the same as those which prevent or reduce fatigue and feelings of eye strain. Luckiesh (**5**) and Luckiesh and Moss (**4**) in their extensive series of investigations, have emphasized mainly rate of involuntary blinking and visibility as criteria of readability. They consider blink-rate particularly to be a highly sensitive measure of readability. The usefulness of any measuring device depends upon its reliability and validity.

Tinker (**21**) found visibility measurements to be highly consistent. The reliability of these measurements, therefore, may be considered adequate. In another study, Tinker (**23**) found a satisfactory reliability for blink rates when the blinks were counted during successive 5- or 10-min. periods of reading. But, according to Tinker (**21**), the validity of visibility measurements as an index of readability is poor except in special situations such as for variations in brightness contrast between print and paper. Adequate information is not available, however, concerning the validity of blink rate as an index of readability.

The present investigation is concerned with the validity of blink frequency as a criterion of readability. The reading of text in roman lower case was compared with the reading of text in all-capitals. These typographical arrangements were chosen for comparison because text in all-capitals is read markedly slower than text in lower case, and because readers dislike very much to read materials printed in all-capitals.

## Procedure

The experiment was carried out in a light laboratory with 10 foot-candles of well diffused illumination. Sixty university students served as readers for two experimental sessions. The

* Grateful acknowledgment is given to the Graduate School, University of Minnesota, for research grant to finance this study.

reading material was placed on an inclined stand at the normal reading distance from the *S*, i.e., about 14 in.

The textual material consisted of Forms I and II of Tinker's [1] Speed of Reading Test. Form II is equivalent to form I. Form I was printed in 10-point lower case Excelsior type with 2-point leading in a 20-pica line width on eggshell paper stock. Form II was typographically the same as Form I except that the text was in all-capitals instead of lower case. There are 450 items of 30 words each in each test form. One word which spoils the meaning is crossed out (indicated orally in this experiment) in the latter part of each item to show that the material has been read and comprehended.

The direct method of observing and counting eye blinks by a skilled *E* was employed (23). Speed of reading in terms of number of items read correctly, as well as number of blinks, was recorded.

On arrival at the laboratory, the *S* was adapted for 10 min. to the illumination used. During this interval he was told that the study was concerned with reading ability, and he was given a practice exercise. At the first sitting the *S* read for 10 min. on Form I (lower case) and, after a five min. rest, read for another 10 min. on Form II (all-capitals). Records were kept of number of eye blinks and speed of reading for each five min. of each 10-min. period. The *S* began with item 1 of each test form.

At the second sitting, after adaptation to the illumination, the *S* read for 10 min. on Form II (all-capitals) and then after a rest, for 10 min. on Form I (lower case). At this sitting the *S* began with item 201 in order to have different text from that read at the first sitting. Again the blinks and speed of reading were recorded for each five min. of the 10-min. period.

## Results

The data for frequency of blinking are listed in Table I. The coefficients of correlation in column 8 represent consistency of blinking from one period to the other. These coefficients, ranging from .83 to .94, represent adequate reliability for the comparisons involved. For each sitting, frequency of blinking while reading all-capitals was compared with frequency of blinking while reading lower case for the first five min., for the second five min., and for 10 min. (5 + 5). Similar comparisons were made for reading done at the second sitting. Then the data for both sittings were combined, yielding a comparison for reading the two kinds of typography for a 20-min. period each.

The mean scores in column 4 reveal that on the average 25 to 30 blinks are made during five min. of reading. Differences (mean for lower case minus mean for all-capitals) and percentage differences are given in columns 6 and 7. In two instances the text for all-capitals was read with slightly more blinks than the text in lower case. In all other comparisons the text in all-capitals was read with slightly less blinks than the text in lower case. The significance of these differences, computed by use of the formula for correlated measures, is given in column 9. No difference is statistically significant in 6 of the 7 comparisons. In the only comparison where the difference approaches significance ($D/PE_{diff.} = 3.93$), the lower case material is read with more blinks than the all-capitals. There is, therefore, no significant difference in blink rate for reading text in all-capitals in

[1] Developed for the National Research Council. It is not yet available for commercial distribution.

comparison with text in lower case. The over-all trend, shown for reading each kind of material for 20 min., is an insignificantly greater number of blinks for reading the text in lower case. If rate of blinking is accepted as a criterion of readability, therefore, one would be forced to conclude that there is no difference between lower case and all-capitals in readability. Before such a conclusion is

TABLE I

INFLUENCE OF TYPE FORM ON FREQUENCY OF BLINKING

Differences are expressed as mean score on Form I minus mean score on Form II of the Tinker Speed of Reading Test. Form I was printed in Excelsior type face, 10-point lower case, in a 20-pica line width with 2-point leading on eggshell paper stock. Form II was typographically the same as Form I, except that it was printed in all-capitals rather than in lower case. $N = 60$ college students.

| Experimental Session | Period | Test Form | Mean | $PE_M$ | Difference Between Means in | | $r$ | $\frac{D}{PE_{diff.}}$ |
|---|---|---|---|---|---|---|---|---|
| | | | | | Blinks | Percent | | |
| (1) | (2) | (3) | (4) | (5) | (6) | (7) | (8) | (9) |
| 1 | 1st 5 min.<br>1st 5 min. | I, l. c.<br>II, all-caps | 25.67<br>26.92 | 1.57<br>1.62 | −1.25 | −4.87 | .83 | 1.33 |
| 1 | 2nd 5 min.<br>2nd 5 min. | I, l. c.<br>II, all-caps | 31.88<br>31.08 | 1.82<br>1.71 | .80 | 2.51 | .92 | 1.13 |
| 1 | 1st 10 min.<br>1st 10 min. | I, l. c.<br>II, all-caps | 57.55<br>58.00 | 3.32<br>3.26 | −.45 | −.78 | .92 | 0.34 |
| 2 | 1st 5 min.<br>1st 5 min. | I, l. c.<br>II, all-caps | 28.63<br>25.02 | 1.67<br>1.62 | 3.61 | 12.61 | .84 | 3.93 |
| 2 | 2nd 5 min.<br>2nd 5 min. | I, l. c.<br>II, all-caps | 29.58<br>28.43 | 1.71<br>1.66 | 1.15 | 3.89 | .84 | 1.19 |
| 2 | 1st 10 min.<br>1st 10 min. | I, l. c.<br>II, all-caps | 58.22<br>53.45 | 3.33<br>3.24 | 4.77 | 8.19 | .86 | 2.78 |
| 1 + 2 | All 20 min.<br>All 20 min. | I, l. c.<br>II, all-caps | 115.77<br>111.45 | 6.40<br>6.25 | 4.32 | 3.73 | .94 | 1.90 |

drawn, however, let us examine the rates of reading for the two kinds of typography.

The basic data for the speed of reading comparisons are given in Table II. In every comparison the text in all-capitals is read slower than the text in lower case. The percent retardation in speed of reading ranges from 9.53 to 19.01. In the over-all comparison (20 min. for each kind of reading) the retardation is approximately 14 percent for the all-capitals material. This, incidentally, is a slightly greater loss than that (about 12 percent) reported by Paterson and Tinker (**8, 14**) in a study involving a relatively short reading period. The critical ratios shown in column 9 show that all the differences are

TABLE II

INFLUENCE OF TYPE FORM ON SPEED OF READING

(See caption of Table I for typographical arrangements)

| Experimental Session | Period | Test Form | Mean | $PE_M$ | Difference Between Means in | | $r$ | $\frac{D}{PE_{diff.}}$ |
|---|---|---|---|---|---|---|---|---|
| | | | | | Paragraphs | Percent | | |
| (1) | (2) | (3) | (4) | (5) | (6) | (7) | (8) | (9) |
| 1 | 1st 5 min. | I, l. c. | 53.52 | .83 | 5.10 | 9.53 | .86 | 11.75 |
| | 1st 5 min. | II, all-caps | 48.42 | .77 | | | | |
| 1 | 2nd 5 min. | I, l. c. | 58.53 | 1.08 | 6.13 | 10.47 | .91 | 13.60 |
| | 2nd 5 min. | II, all-caps | 52.40 | 1.01 | | | | |
| 1 | 1st 10 min. | I, l. c. | 112.05 | 1.87 | 11.23 | 10.02 | .92 | 14.78 |
| | 1st 10 min. | II, all-caps | 100.82 | 1.75 | | | | |
| 2 | 1st 5 min. | I, l. c. | 58.56 | 1.12 | 11.13 | 19.01 | .86 | 19.05 |
| | 1st 5 min. | II, all-caps | 47.43 | .82 | | | | |
| 2 | 2nd 5 min. | I, l. c. | 61.85 | 1.29 | 9.93 | 16.05 | .89 | 16.23 |
| | 2nd 5 min. | II, all-caps | 51.92 | .96 | | | | |
| 2 | 1st 10 min. | I, l. c. | 120.42 | 2.38 | 21.07 | 17.50 | .91 | 19.47 |
| | 1st 10 min. | II, all-caps | 99.35 | 1.76 | | | | |
| 1 + 2 | All 20 min. | I, l. c. | 232.47 | 4.03 | 32.30 | 13.89 | .96 | 27.21 |
| | All 20 min. | II, all-caps | 200.17 | 3.43 | | | | |

statistically significant at the one percent level. These results indicate that the readability of text printed in all-capitals is significantly poorer than the readability of the lower case material. This reduced readability of text in all-capitals in comparison with lower case is undoubtedly due, not to reduced visibility in the former, but to lack of distinctive word forms in the all-capital printing. See below for further discussion of this factor.

## DISCUSSION

Two markedly different results appear in this experiment. If rate of blinking is accepted as a criterion of readability, there is no difference in the readability of all-capital and lower case materials. But the text in all-capitals retards reading by an amount seldom exceeded in non-optimal typographical arrangements. This fact is revealed not only in this study but also in earlier investigations (**8**). Furthermore, 90 percent of readers consider lower case to be more readable than all-capitals. Incidentally, about the same proportion of readers judge the lower case to be more pleasing (**18**). If reading rate is markedly retarded and readers judge the material to be less

readable, it would seem reasonable to conclude that text printed in all-capitals is less readable than text in lower case. There are several factors which support such a conclusion. When eye movements are photographed while reading the two kinds of material, Tinker and Paterson (**16**) found that reading text in all-capitals required significantly more fixations and longer pause durations than text in lower case. These patterns of eye movements indicate poorer readability for the all-capital material. In a long series of studies, Tinker and Paterson (**7, 9, 10, 11, 12, 16, 17**) have found that ineffectual eye movement patterns coincide with retarded rate of reading in indicating poorer readability for typography that varied from the optimal arrangement. Luckiesh and Moss (**4**, p. 255) suggest that variations in oculomotor patterns reveal differences in readability.

Another line of supporting evidence comes from studies of newspaper headlines printed in all-capitals and in lower case. Working with single-column headlines, Breland and Breland (**2**) found, with a short exposure technique, that there was 18.9 percent of loss in readability in reading the headlines set in all-capitals in comparison with lower case. In a somewhat similar study of banner headlines exposed six feet from the reader, Paterson and Tinker (**13**) discovered that the readability of the headlines in lower case was 5.3 percent greater than for those in all-capitals.

It is well known that 'total word form' or the characteristic shape of words is an important factor in promoting ready perception of words. These word forms are far more characteristic when words are printed in lower case than when they are printed in all-capitals. Tinker (**15**) found that the total word form is more potent in the perception of words in lower case than in all-capitals. In the all-capital material, perception seemed to occur largely letter by letter. As noted by Webster and Tinker (**24**), the influence of word form in promoting readability is especially important in reading connected discourse.

It may be inferred from the data cited above and from the analysis presented that text in all-capitals is less readable than text in lower case. Nevertheless, rate of blinking, which is emphasized by Luckiesh and Moss (**4**) as a criterion, reveals no difference in readability for the two typographical arrangements. In terms of the data presented above, therefore, rate of blinking must be rejected as a criterion of readability. Before accepting this conclusion it is desirable to make a critical analysis of other findings and discussions, since one might argue that, although no differences in blink rate are found in this study, the technique might well be valid on other situations.

Tinker has already raised the question of validity of the blink

rate technique as a measure of ease of seeing in two book reviews (**19**, **22**) and in a discussion (**20**). There are other relevant experimental data. After a searching analysis of their own findings and of data from other studies, McFarland, Holway and Hurvich (**6**) state that a high blink-rate need mean neither an increase in visual fatigue nor an increase in difficulty of seeing, and conclude that "The rate of blinking can hardly be considered as a valid index of visual fatigue." Bitterman (**1**) obtained work output and rate of blinking records while *S*s read and responded (by checking) to a test requiring exact visual discrimination. One form of the material was in 10-point type and the second form in 6-point. He found that the work output was significantly greater for reading and responding to the large type, but that there was no significant difference between the two in rate of blinking. The insignificant difference present represented fewer blinks for working with the smaller type. In the same study, Bitterman determined rate of blinking for reading 10-point type under three and under 91 foot-candles of light. Frequency of blinking was slightly greater for reading under the bright (91 foot-candles) light but the trends of the differences were not significant. The author concludes that rate of blinking can have little value as an index of visual fatigue or effort (i.e., ease of visual work) and that the results, therefore, do not support the conclusions of Luckiesh and Moss. In other words, rate of blinking cannot be accepted as a criterion of readability and ease of seeing.

It is interesting and highly significant that no independent scientist, whether he is working at Harvard, Columbia, Cornell, or the University of Minnesota, is able to find the same trends as Luckiesh and his co-workers when he examines frequency of blinking as an index of readability or of ease of seeing Eames (**3**) criticises Luckiesh and Moss (**4**) for using relatively few *S*s in their experiments and for employing 'test wise' *S*s. As pointed out by Eames, "People who take tests repeatedly in a given field gradually learn what is expected of them" and are unintentionally influenced by this. Results obtained under such conditions cannot be representative of the reactions of the general population. The low probable errors reported by Luckiesh and Moss are what might be expected from *S*s who are thoroughly 'test wise.'

Although more checking on the validity of the blink technique as a measure of readability and of ease of seeing is desirable, there is enough evidence at hand to suggest that the findings of Luckiesh and his co-workers are due to an artifact which has led to fallacious conclusions. The condition which could produce this is the use of 'test wise' *S*s. Continued failure to substantiate the rate of blinking as a criterion of ease of seeing and of readability must necessarily

lead to its rejection as an adequate technique in studies of visual work. This would mean that much of the evidence put forward by Luckiesh and by Luckiesh and Moss to support recommendations for desirable practice in the field of illumination and typographical arrangements is worthless and that the recommendations, therefore, would be without sound foundation. In any case, since there is grave doubt concerning the validity of the blink technique as a criterion of ease of seeing and of readability of print, the results obtained by Luckiesh and his co-workers and the recommendations derived from the results should not be accepted at face value pending further investigation.

## Summary and Conclusions

1. The purpose of the present investigation is to check the validity of frequency of blinking as a criterion of readability.

2. Frequency of blinking and rate of reading records were obtained while reading text printed in lower case and in all-capitals.

3. Sixty university students served as readers.

4. There was no significant difference in rate of blinking while reading text in all-capitals in comparison with lower case.

5. In every comparison made, text in all-capitals was read significantly slower than text in lower case. The percent retardation in speed of reading ranged from 9.53 to 19.01.

6. Analysis of the data in this experiment plus consideration of evidence from other studies indicates that the frequency of blinking is an unsatisfactory criterion of readability of print.

7. The data furnish additional evidence that rate of reading is a fairly adequate index of readability. This does not mean that other measures may not be employed to supplement rate of reading in situations where the validity of such measures has been established. Limitations of rate of reading as a criterion of ease of seeing have been recognized by Paterson and Tinker (**8**).

8. The failure of independent investigators to confirm the trends reported by Luckiesh and his co-workers suggests that the validity of frequency of blinking as a criterion of readability and of ease of seeing is in doubt.

(Manuscript received January 7, 1946)

### References

1. Bitterman, M. E. Heart rate and frequency of blinking as indices of visual efficiency. *J. exp. Psychol.*, 1945, **35**, 279–292.
2. Breland, K., & Breland, M. K. Legibility of newspaper headlines printed in capitals and in lower case. *J. appl. Psychol.*, 1944, **28**, 117–120.
3. Eames, T. H. Review of *Reading as a visual task* by M. Luckiesh and F. K. Moss. *The Columbia Optometrist*, 1942, **16**, 7.

4. Luckiesh, M., & Moss, F. K. *Reading as a visual task.* New York: D. Van Nostrand Co., 1942, pp. 428.
5. Luckiesh, M. *Light, vision and seeing.* New York: D. Van Nostrand Co., 1944, pp. 323.
6. McFarland, R. A., Holway, A. H., & Hurvich, L. M. *Studies of visual fatigue.* Boston: Graduate School of Business Administration, Harvard Univ. (Soldiers Field), 1942, pp. 255.
7. Paterson, D. G., & Tinker, M. A. Influence of line width on eye movements. *J. exp. Psychol.*, 1940, **27**, 572–577.
8. Paterson, D. G., & Tinker, M. A. *How to make type readable.* New York: Harper & Brothers (obtainable from the authors), 1940, pp. 209.
9. Paterson, D. G., & Tinker, M. A. Influence of size of type on eye movements. *J. appl. Psychol.*, 1942, **26**, 227–230.
10. Paterson, D. G., & Tinker, M. A. Influence of line width on eye movements for six-point type. *J. educ. Psychol.*, 1942, **33**, 552–554.
11. Paterson, D. G., & Tinker, M. A. Eye movements in reading type sizes in optimal line widths. *J. educ. Psychol.*, 1943, **34**, 547–551.
12. Paterson, D. G., & Tinker, M. A. Eye movements in reading optimal and non-optimal typography. *J. exp. Psychol.*, 1944, **34**, 80–83.
13. Paterson, D. G., & Tinker, M. A. Readability of newspaper headlines printed in capitals and in lower case. *J. appl. Psychol.*, 1946, **30**, 161–168.
14. Tinker, M. A., & Paterson, D. G. Influence of type form on speed of reading. *J. appl. Psychol.*, 1928, **12**, 359–368.
15. Tinker, M. A. The influence of form of type on the perception of words. *J. appl. Psychol.*, 1932, **16**, 167–174.
16. Tinker, M. A., & Paterson, D. G. Influence of type form on eye movements. *J. exp. Psychol.*, 1939, **25**, 528–531.
17. Tinker, M. A., & Paterson, D. G. Eye movements in reading a modern type face and Old English. *Amer. J. Psychol.*, 1941, **54**, 113–115.
18. Tinker, M. A., & Paterson, D. G. Reader preferences and typography. *J. appl. Psychol.*, 1942, **26**, 38–40.
19. Tinker, M. A. Review of *Reading as a visual task* by M. Luckiesh and F. K. Moss. *J. appl. Psychol.*, 1943, **27**, 116–118.
20. Tinker, M. A. A reply to Dr. Luckiesh. *J. appl. Psychol.*, 1943, **27**, 469–472.
21. Tinker, M. A. Criteria for determining the readability of type faces. *J. educ. Psychol.*, 1944, **35**, 385–396.
22. Tinker, M. A. Review of *Light, vision and seeing* by M. Luckiesh. *J. appl. Psychol.*, 1945, **29**, 252–253.
23. Tinker, M. A. Reliability of blinking frequency employed as a measure of readability. *J. exp. Psychol.*, 1945, **35**, 418–424.
24. Webster, H. A., & Tinker, M. A. The influence of type face on the legibility of print. *J. appl. Psychol.*, 1935, **19**, 43–52.

# THE EFFECTS OF CERTAIN PRE-TRAINING PROCEDURES UPON MAZE PERFORMANCE AND THEIR SIGNIFICANCE FOR THE CONCEPT OF LATENT LEARNING

BY HARRY W. KARN AND JAMES M. PORTER, JR.

*Carnegie Institute of Technology*

## Introduction

It is a generally accepted principle of maze learning that an animal without food reward does not exhibit error reduction in successive trials comparable to that of an animal fed at the end of the maze. However, despite a demonstration of little or no error reduction during non-rewarded runs, it is reasonable to assume that the behavior of the animal is in some way modified as the result of this experience with the maze enclosure. One way of testing this assumption is to study the progress made by a group of animals in acquiring a maze habit after they have explored the apparatus without food reinforcement. Varied in a number of ways this is essentially the procedure followed in latent learning experiments. Since the present investigation is concerned with the effects of certain pre-training procedures on subsequent learning in a maze, a consideration of the main features of the latent learning experiment is pertinent.

As Reynolds (**4**) recently pointed out, it is possible to distinguish two main types of latent learning experiments among those reported in the literature. The first involves subjecting a group of animals to a series of non-rewarded trials followed by a series of rewarded trials. A comparison is made of the error reduction between the first and second rewarded trials and the greatest trial-to-trial decrease shown by a group of animals which had been fed from the initial trial. Blodgett (**1**) appears to have been the first investigator to use this method, and he reports a greater error reduction for the first group. These findings are interpreted by Blodgett as due to latent learning during the non-rewarded runs.

The second type of latent learning experiment is exemplified in a study by Haney (**2**) in which a group of animals was allowed to wander in a maze for a period of time before the beginning of food-reinforced training runs. During subsequent rewarded trials this group made fewer errors than a control group of animals learning the experimental

maze for the first time. This superiority was attributed by Haney to the operation of latent learning.

Herb (3), Tolman (5), and Tolman and Honzik (6, 7) have also reported demonstrations of latent learning.

Although Tolman (5) presents an account of the hypothetical factors which operate to modify the animal's behavior during the period when latent learning is allegedly taking place, there has apparently been no report of an attempt to isolate these factors experimentally. To the present investigators such an analysis appears necessary, and is essentially one that must be undertaken through a study of the influences of various pre-training procedures upon subsequent learning. Accordingly, the present report is concerned with an experimental investigation of the effect on maze performance of (a) habituation to handling, (b) familiarity with the maze, (c) habituation to confinement, and (d) goal orientation.

## Method

Four experimental groups and a control group of white rats were used as *Ss*. The maze training was carried out in a modified form of the Dashiell maze (Fig. 1). This apparatus was used because of its adaptability to analysis of the orientation factor in maze learning.

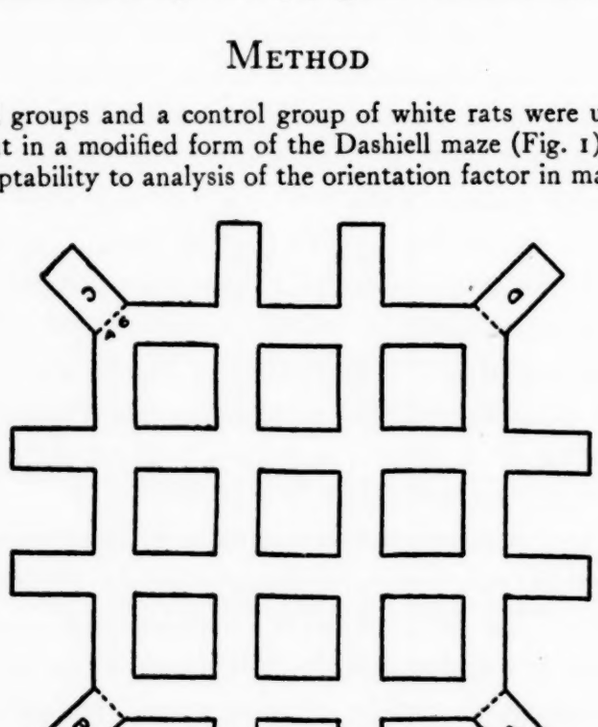


Fig. 1. Ground plan of the modified Dashiell maze used in this investigation

The pre-training and the training procedures employed with each group of animals were as follows:

Group 1. The control group (14 animals) received no pre-training attention and their rewarded training was identical with that of the other groups.

Group 1A. Each of the 12 animals in this group received individual handling once a day for six consecutive days. This handling consisted in each animal being transferred by hand from the living cage to another cage and later back to the living cage again. The living cage contained food and water at all times during the six-day period. Following handling on the sixth day the food was removed from the cage and the animals fed thereafter once a day for 30 min. On the seventh day, each animal was placed in starting compartment A of the maze and released after

one minute. When the animal arrived at goal compartment C, a door was quietly closed behind it and the animal was allowed to feed for one min. on bread soaked in milk, after which it was returned to the living cage where it was allowed to feed for 30 min. This food-rewarded training was continued at the rate of one trial a day until a criterion of two successive errorless runs was attained.

Group 2. The 12 animals in this group were handled individually for six days under the same conditions described for the members of Group 1A. On the seventh day and for the following four days each of these animals was given one trial per day under the following conditions. After being detained in starting box A for one min. the animal was released into the maze. When it entered goal box C, the door was closed behind it and the animal was detained without food for one min. The animal was then removed from the goal box and placed in an individual detention cage for about one hour. A different detention cage was used each day. At the end of the detention period the animals were returned to their living cages and fed their daily ration of food. Following five days of this kind of non-rewarded training, the animals were given rewarded trials at the rate of one trial per day until the criterion of two successive errorless runs was attained.

Group 3. The 13 animals of this group were handled individually for six days in the same manner as Groups 1A and 2. Each animal was then paired by chance with a member of Group 2 and subjected to the following procedure. After being placed in a starting box and detained for one min., the animal was released into the maze and allowed to wander for a period of time. This period of time was equal to that required by its partner in Group 2 to reach goal box C after being released from starting box A. At the end of this period the animal was picked up wherever it happened to be in the maze and detained for an hour in an individual detention cage. The animals in this group were never released into the maze from the same starting box on two successive days and never more than twice from any one starting box. With the Dashiell maze it is possible to use either compartment A, B, C, or D as a starting compartment. Following these five exploratory trials each member of Group 3 was given rewarded training in the same manner as the members of the other groups.

Group 4. The 14 animals of this group were handled individually in the manner described for previous groups. Each animal was then paired with a member of Group 2. On the subsequent five days the pre-training procedure for members of this group was as follows. Each animal was carried into the experimental room and placed in a box similar to the starting box, but not a part of the maze. The animal was detained in this box for the same length of time that its partner had spent in the maze proper. At the end of this period the animal was placed in an individual detention cage for at least one hour and then removed to its living cage and fed its daily ration of food. Rewarded training in the maze was identical with that described for all preceding groups.

The procedures which have been described were evolved in order to determine the possible facilitating effect on subsequent rewarded training of one or more of the following factors: habituation to handling, the establishment of a goal orientation within the maze, familiarity with the maze without goal orientation, and confinement within an enclosure. Thus, in the case of the animals in Group 2 all these factors were operative; with Group 3 the goal orientation factor was absent but all other factors operative; with Group 4 the only possible facilitating factors were handling and habituation to confinement within an enclosure. The only possible facilitating factor with the members of Group 1A would be handling similar to that received by all other groups.

By a comparative study of the performance of the various groups it appeared plausible that information might be obtained regarding the factors which operate to make latent learning manifest when food reward is introduced. An identification of these factors might conceivably reveal behavior changes that are not susceptible to measurement in terms of error or time scores. If this should be true, there would seem to be justification for questioning the validity of any concept that envisages learning to be latent until it can be measured by the traditional error-time score technique. Blodgett *et al.* apparently inferred that their animals learned during the exploration period, but that this learning was in no way utilized until a food or water incentive was introduced into the maze.

## Results

The median number of trials required to learn the maze to the criterion of two successive errorless runs for each group is shown in

Table I. The reliabilities of the differences between the medians for each group are also presented.

TABLE I

MEDIAN NUMBER OF REWARDED TRIALS TO LEARN THE DASHIELL MAZE

| Group No. | 1 | 1A | 2 | 3 | 4 |
|---|---|---|---|---|---|
| Median | 10.25 | 8.66 | 3.83 | 7.25 | 8.25 |
| $SE_{mdn}$ | 1.19 | .80 | 1.36 | 1.17 | 1.13 |

CRITICAL RATIOS OF THE DIFFERENCES BETWEEN THE MEDIANS

| Group No. | 1A | 2 | 3 | 4 |
|---|---|---|---|---|
| 1 | 1.10 | 3.54 | 1.85 | 1.22 |
| 1A | | 3.05 | .99 | .29 |
| 2 | | | 1.91 | 2.49 |
| 3 | | | | .61 |

An arrangement of the data in terms of complexity of pre-training procedures and their relevance to the problem, i.e., mastering the correct pathway from A to C, indicates the following trends:

1. Animals which received handling plus familiarity in the maze plus goal orientation (without food-reinforcement) show the most rapid progress of all groups in reaching the criterion of learning when food is introduced into the goal box.

2. Animals which received handling and familiarity with the maze, but without goal orientation, required a greater number of rewarded trials to attain the criterion of learning than the group of animals for whom the added factor of goal orientation was operative.

3. Animals which received handling plus familiarity with an enclosure (essentially identical with the starting compartment), but not maze familiarity nor goal orientation, required a greater number of rewarded trials to attain the learning criterion than groups which had been subjected to either or both of these latter pre-training procedures.

4. Animals which received only handling prior to rewarded training required the greatest number of rewarded trials to attain the learning criterion of all the groups subjected to pre-training procedures.

5. A control group of animals introduced into the maze without any pre-training made slower progress in attaining the criterion of learning than animals subjected to any one or more of the pre-training procedures employed in this investigation.

These trends are depicted graphically in Fig. 2, which shows the learning curves expressed as Vincent scores during rewarded trials for all groups.

Time scores for the rewarded trials stand in essentially the same relationship as error scores, and therefore are not presented.

After the completion of the study of the effect of pre-training procedures upon subsequent learning of the Dashiell maze, all groups were trained in a standard six-unit T maze. An analysis of the results of this test revealed no significant differences in learning ability for any one group of animals. Therefore, the differnces obtained in the main experiment would appear to be due to the effect

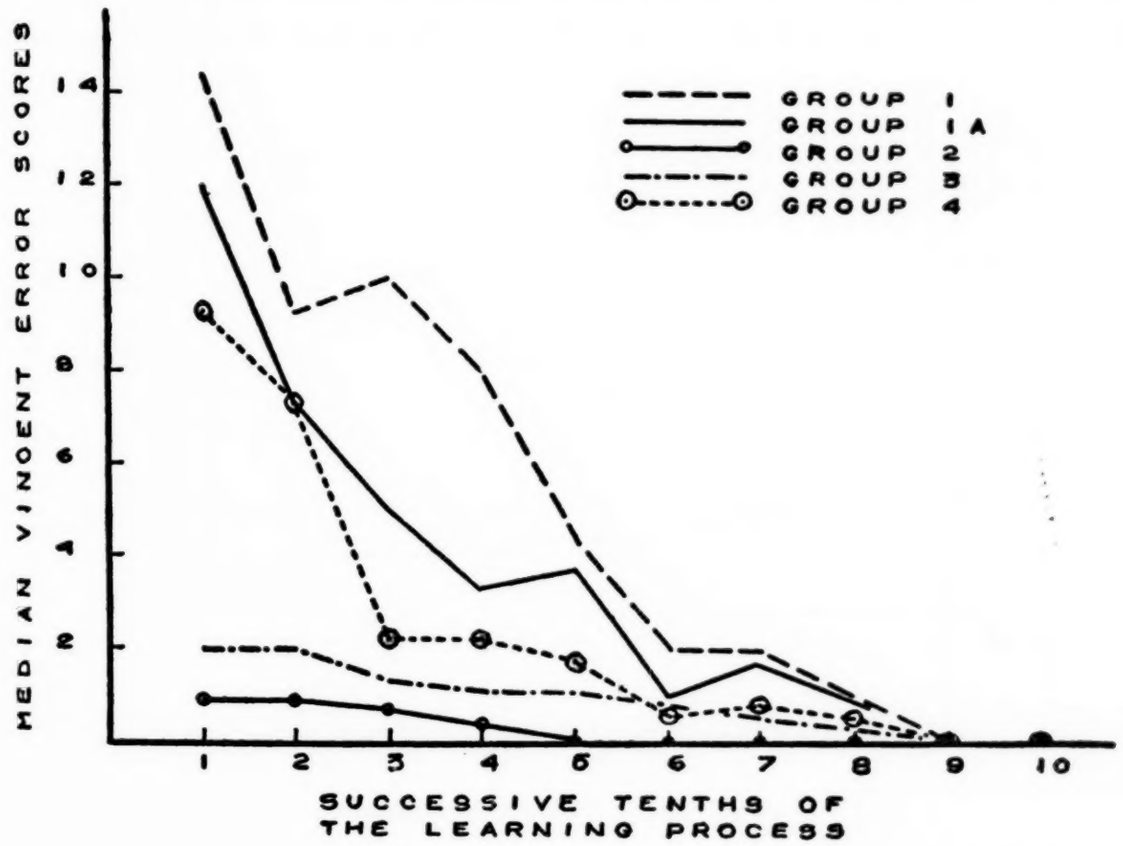


FIG. 2. Learning curves for control and experimental groups of animals during rewarded training

of the various pre-training procedures and not the result of differences in learning ability on the part of the groups.

The pre-training performance of Group 2 presents several points of interest which merit further consideration, particularly since it was only with this group that statistically significant differences were obtained; essentially the procedure employed with this group was that used by investigators of latent learning. It will be recalled that Group 2 had been subjected to the most extensive pre-training procedure, including handling, maze familiarity, and goal orientation. If the animal's performance during the five non-rewarded trials is analyzed in the same manner as its later performance, i.e., any deviation from a goal pointing path, and each unit so traversed, is

counted as one error, then evidence may be sought for partial mastery of the maze previous to the introduction of food in compartment C. Fig. 3, which depicts the performance of Group 2 during the five-non-rewarded trials shows little decrease in errors during the first four trials and, in fact, a large increase on the fifth trial.

The results of this investigation have undoubtedly been influenced by a factor which thus far has not been mentioned. A large number of animals in Group 2 manifested what has been generally termed conflict behavior. This behavior appeared to result from a hesitancy on the part of the animals to enter the goal box, and, at the same time,

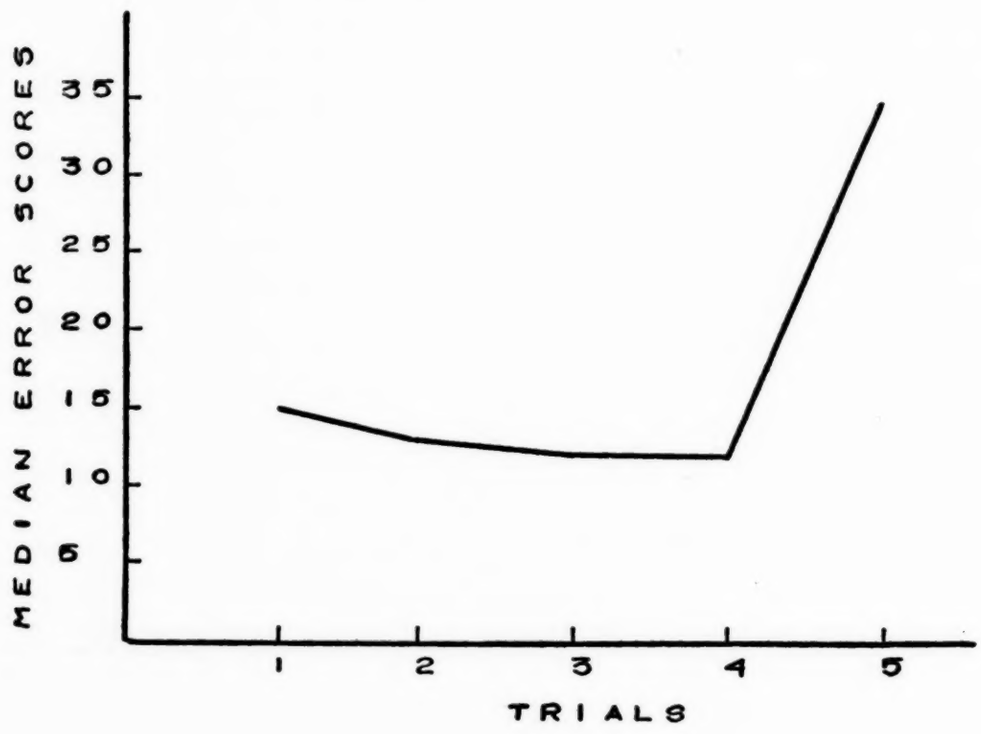


FIG. 3. Learning curve for Group 2 during non-rewarded trials

a tendency to enter the box. This hesitancy was apparently engendered by the closing of the goal compartment door behind the animal. The apparent conflict manifested itself in the animal's partial entry into the goal box, followed by turning away and going to other parts of the maze before finally entering the goal box. This, of course, served to raise the error score for that trial. The disruptive effects of the apparent conflict seemed to manifest themselves for the first time on the second or third pre-training trials.

In an attempt to ascertain what the animals' performance would have been during the first five trials had the experimental situation not engendered conflict the data have been treated as follows. Since the disruptive effects were first observed in section 46 of the maze

(see Fig. 1) a determination has been made of the number of errors made by each animal before reaching section 46 for the first time. This is at very best a doubtful procedure, but the results are interesting. When the number of errors made before section 46 is entered are plotted, the median number of errors per trial is indicated by the curve in Fig. 4, which shows considerable learning during the period of no food reward.

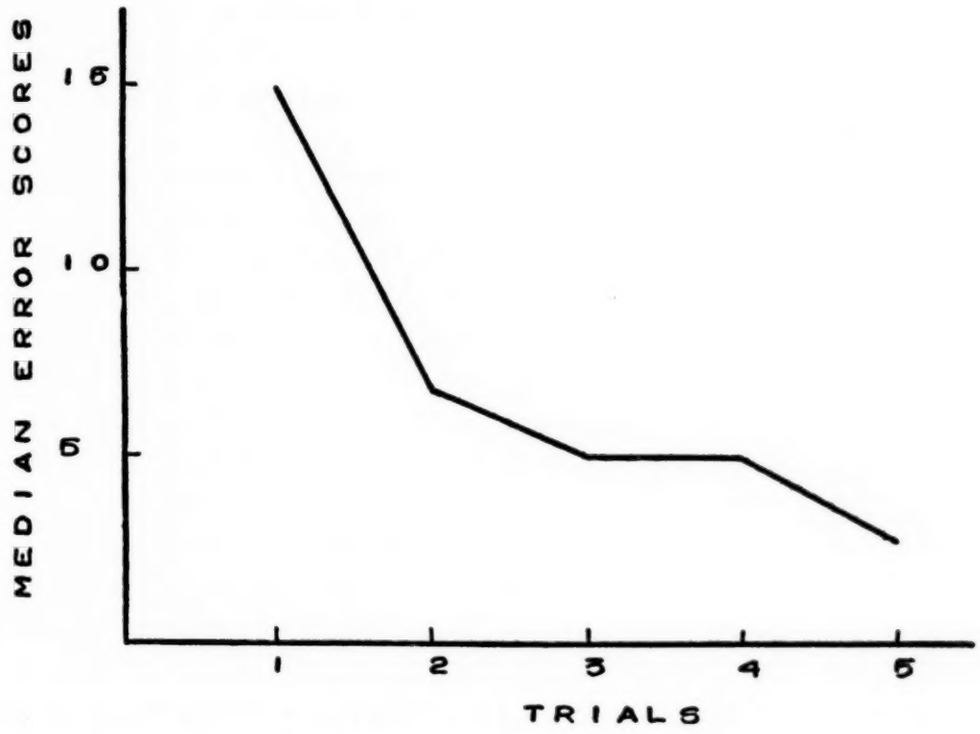


FIG. 4. Learning curve for Group 2 during non-rewarded trials corrected for conflict

## DISCUSSION

To the writers, the overall results of the present investigation suggest that pre-training maze experience results in negative adaptation or the minimizing of emotional and other inhibitory factors, such as the initial tendency to explore, which in the conventional maze experiment serve to retard the acquisition of the maze habit in terms of the early elimination of blind alleys and speed of traversing the pathway between starting box and goal box. The present experiment also indicates that pre-training maze experience involves another factor, in the form of an orientation toward the goal which serves to facilitate subsequent learning when food or some other reward is placed in the goal box.

When at least some of the factors mentioned in the preceding paragraph are isolated and experimentally investigated, evidences are revealed which show that those animals will profit most whose

pre-training experiences include the operation of the greatest number of these factors. Thus, animals which receive handling plus maze familiarity plus goal orientation show more rapid progress in acquiring the maze habit than do those groups whose previous training does not permit the operation of all of these factors. That factors other than goal orientation also provide facilitory effects is shown by the superior performance of groups subjected to handling, confinement, and maze familiarity, without goal orientation, over the control group which had no pre-training experience. It would seem to be a warranted assumption that some or all of the factors studied in the present experiment are operative in the Blodgett and Haney types of latent learning experiments.

The results of previously reported latent learning experiments have generally been interpreted as due to the building up of a set of conditions within the animal which do not manifest themselves until a food goal is introduced (a view hereinafter termed the latent learning hypothesis). The present writers favor the hypothesis that the modifications of the animals' behavior which occur before the introduction of a food reward are apparently held in abeyance or remain latent only because of the failure or inability of the experimenter to record them. In other words, a more refined recording technique, possibly involving the measurement of respiration, pulse rate, or other emotional indices, would reveal objective evidence of progressive adjustment to the maze situation prior to the introduction of food reward. In keeping with this predictive interpretation it is interesting to note that Reynolds (**4**) suggests a refinement in measurement in terms of depth of penetration of blind alleys during the so-called period of latent learning. Indeed, in a repetition of Blodgett's earlier study of latent learning in the maze, Reynolds claims to have found a gradual decrease in depth of blind-alley penetration during the period of no food reward, and on the basis of this analysis finds no evidence for truly latent learning.

An additional point which appears to weaken the latent learning hypothesis is the analysis of conflict behavior in one of the groups of animals in the present study. It will be recalled that evidence of learning during the non-rewarded trials is apparent with that group which had been subjected to the most extensive pre-training experience. Similar evidence for objective signs of learning during the non-reward period is present in the form of decreases in blind alley entrances in the maze studies of Herb (**3**) and Tolman and Honzik (**6**, **7**). Reynolds (**4**) emphasizes these data, and concludes that it is incorrect to say that no 'adding up' (Tolman, **5**) occurs in the non-rewarded period of the commonly reported latent learning experiment.

## Summary

Four groups of white rats were subjected to certain pre-training procedures, and the effect upon subsequent learning of a modified Dashiell maze was studied. A control group learned the maze without being subjected to pre-training procedures. The pre-training procedures were designed to give the experimental groups respectively: (1) habituation to handling; (2) familiarity with the maze; (3) familiarity with the maze and goal orientation; and (4) habituation to being detained in an enclosure similar to the starting compartment in the maze.

During subsequent rewarded training the maze was mastered most rapidly by that experimental group whose pretraining was most extensive, involving specifically, handling, maze familiarity, and goal orientation. No statistically significant differences were found between the learning scores of the other experimental groups, although there is an indication of superior performance on the part of all groups that had been subjected to one or more pre-training procedures.

It is suggested that those factors which operated in the present study to produce facilitory effects upon rewarded training were also operative in previously reported latent learning experiments. The hypothesis is advanced that during the period of no food reward there is a resolution of incompatible response tendencies which tend to inhibit time and error reduction but which are nevertheless an essential part of the animals' learned reactions to the maze situation. In some maze situations the subsequent rewarded training may be further enhanced by the establishment of a goal orientation. These facilitory tendencies are objectively verifiable if a sufficiently refined technique of measurement is utilized.


(Manuscript received August 1, 1946 for immediate publication)

## References


1. Blodgett, H. C. The effect of the introduction of reward upon the maze performance of rats. *Univ. Calif. Publ. Psychol.*, 1928, **4**, 114–134.
2. Haney, G. W. The effect of familiarity on maze performance on albino rats. *Univ. Calif. Publ. Psychol.*, 1928, **4**, 319–333.
3. Herb, F. H. Latent learning—non-reward followed by food in blinds. *J. comp. Psychol.*, 1940, **29**, 247–256.
4. Reynolds, B. A repetition of the Blodgett experiment on 'latent learning.' *J. exp. Psychol.*, 1945, **35**, 504–516.
5. Tolman, E. C. The determiners of behavior at a choice point. *Psychol. Rev.*, 1938, **45**, 1–41.
6. Tolman, E. C., & Honzik, C. H. Degrees of hunger, reward, and non-reward, and maze learning in rats. *Univ. Calif. Publ. Psychol.*, 1928, **4**, 31–55.
7. Tolman, E. C., & Honzik, C. H. Introduction and removal of reward and maze performance in rats. *Univ. Calif. Publ. Psychol.*, 1928, **4**, 241–256.